PERGAMON Computers and Mathematics with Applications 40 (2000) 1315–1347

An International Journal
computers & mathematics
with applications

www.elsevier.nl/locate/camwa

Optimal and Near Optimal Quantization of Integrable Functions

T. KÄMPKE
Forschungsinstitut für Anwendungsorientierte Wissensverarbeitung (FAW)
Helmholtzstr. 16, 89081 Ulm, Germany
kaempke@faw.uni-ulm.de

(Received and accepted July 1999)

Abstract—A continuous function is approximated by step functions with given number of break points. The approximations are nonlinear. Local as well as global optimization strategies based on fixed-point analysis are integrated with branch and bound techniques for multivariate Lipschitz minimization. Branching strategies are based on the structure of attracting domains. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords—Fixed-point theory, Global optimization, Lipschitz analysis, Vector quantization.

1. INTRODUCTION

Quantization of the set $\mathbb{R}$—also called scalar quantization—amounts to defining a real-valued step function s on $\mathbb{R}$ and replacing every $x \in \mathbb{R}$ by $s(x)$. Vector quantization is the obvious generalization to the multidimensional Euclidean space $\mathbb{R}^N$ with integer N. Quantization is always guided by some real-valued function f which must be approximated by the step functions. Such problems have been given consideration by many authors including the general work [1–4] and for example the recent special work [5]. Only $N = 1$ dimension will be treated here.

The basic difference between the current type of quantization and standard vector quantization is the distance measure between f and the step functions which are required to have a bounded number of jumps. While the standard criterion $\|\cdot\|_{\text{quant}}$ (see below) tends to assign break points where f is large, the present criterion $\|\cdot\|_p$ (standard p-norm) tends to assign break points where the change of f is large, i.e., where $|f'|$ is large if f is differentiable. The two criteria yield nonlinear optimization problems, because both levels and break points have to be determined simultaneously. The use of the p-norm as criterion of fit is motivated by various segmentation problems in computer vision such as path following problems in computer tomography [6,7] and by work on adaptively encoding various classes of images [8]. Quantization functions are chosen to be simpler than those of Mumford [3] and Blake and Zisserman [9]. Consequently, the same holds for the measure of fit which is an "energy" function in other cases. The present approach allows to get rid of trade offs between fit to the original signal and variability within the approximating signal or even more involved trade offs such as weighting coefficients in the measure of fit.

0898-1221/00/$ - see front matter Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
PII: S0898-1221(00)00241-8

Many constructions from approximation theory as, e.g., best approximations from linear subspaces are only partially applicable to quantization. The reason is that the set of step functions with bounded number of jumps is not a linear space due to break points being variable. The reverse of the stated problem has also been considered. Assume a histogram is to be approximated in a properly defined square metric by some smooth curve with prescribed interpolating points. Appropriate spline fitting algorithms are known for this problem and local minima usually are global ones; see for example [10].

The ultimate aim behind this investigation is to obtain procedures in signal analysis which are ideally free of threshold values. This is related to [3,2,11,12]. Thresholds appear to be inevitable for example in stopping criteria. However, here we do not require to specify parameters such as explicit acceptance levels, window sizes of convolutions, etc. The avoidance of thresholds serves as a guideline for algorithmic developments. As a consequence, we focus on variational principles to some degree and mainly on optimization methods.

The remainder of this paper is organized as follows. In Section 2, we formally define the quantization problem and give some background results. In Section 3, we consider partially known components of a quantization and give corresponding closed form solutions based on local variation techniques. Either the partition or the levels of the step functions are held fixed. The further requires integration (in case $1 \leq p < \infty$) while the latter requires also function inversion. Both operations applied in an alternating mode allow both partitions and levels to vary. A fixed-point formulation of the problem is presented in Section 4. This yields a sequence of successive approximations to converge towards the optimal quantization over a compact domain $[a, b]$ if initial values are suitably chosen. The quantization problem turns out to be one of global minimization with all local optima being characterized in terms of fixed-point behaviour for a wide class of functions called normal. A Lipschitz argument then transforms one-dimensional quantization problems into multivariate Lipschitz optimization problems over compact and convex domains for which a globally convergent branch and bound procedure is presented. Deformation methods which tend to avoid certain local minima will be considered in Section 5. In the final Section 6, we add some stability results. Thereby the effect of varying the signal f on the quantization function is shown to be smooth.

Algorithms are treated in a conceptual manner. The given function f, integrals thereof and other derived quantities, need not be given in closed form but in the sense of an oracle: values are supposed to be known once the arguments are specified. Numerical issues raised by computations of integrals, termination criteria, error bounds, etc., are not considered here.

In the sequel, ∎ marks the end of an argument, $x^\top$ denotes the transpose of (column) vector x, $\mathrm{int}(A)$ and $\partial(A)$ denote the interior and boundary of set A, respectively, and $:=$ and $:\Longleftrightarrow$ denote definitions.

2. QUANTIZATION

Let $f \in C(D) \cap L^p(D)$ with $p \in [1, \infty]$ and $D \in \{\mathbb{R}, \mathbb{R}_{\geq}, [a, b]\}$ be given. Step functions are considered which result from a finite partition of $\mathbb{R}$ given by $x_1 < \cdots < x_n$ and values $y_1, \ldots, y_{n-1} \in \mathbb{R}$, where $n \geq 2$. The set of such step functions is denoted by S_n and each individual function is denoted by

$$s_n = \sum_{i=1}^{n-1} y_i \, 1_{(x_i, x_{i+1})},$$

where 1_A is the indicator function of set A. Function s_n has up to n jumps and $s_n(x) = 0$ for $x < x_1$ and for $x > x_n$. Trivially, the intervals $(-\infty, x_1)$ and (x_n, ∞) cannot be taken into consideration for $p < \infty$ as the corresponding indicator functions are not integrable. Since f is continuous, s_n may be arbitrarily defined at $x_1, \ldots, x_n$ to be right or left continuous without effecting the measures of fit. Each step function is required to be right or left continuous everywhere so that a step function from S_n has at most n values.

The p-norm $\|\cdot\|_p$ on D is given for $p \in [1,\infty]$ as usual by

$$\|f\|_p = \begin{cases} \sqrt[p]{\int_D |f(x)|^p\,dx}, & 1 \le p < \infty, \\ \sup_{x\in D} |f(x)|, & p = \infty, \end{cases}$$

with $\|f\|_2 = \|f\|$.

DEFINITION 1. *An optimal quantization of an integrable function f with respect to the p-norm and fixed number n of jumps is defined by a solution $s_n^0 = \sum_{i=1}^{n-1} y_i^0 1_{(x_i^0, x_{i+1}^0)}$ of the optimization problem*

$$\min_{s_n\in S_n} \|f - s_n\|_p = \left\|f - s_n^0\right\|_p.$$

Quantization is related to appproximations according to the Mumford-Shah energy function $\int_D \|f-g\|^2 d\lambda^2 + \int_{D-K} \|\nabla g\|^2\,d\lambda^2 + r\cdot l(K) =: \varepsilon(g)$ [2] for functions $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^2$. There, D_i are disjoint open and connected sets separated by curves whose total length in $\mathrm{int}(D)$ equals $l(K)$ with $D = \cup_i D_i \cup K$. The approximating function g is supposed to have small variation within each D_i and it is allowd to have a large variation across ∂D_i. The constant r is an external weighting factor. The case $\nabla g := 0$ on $D - K$ and the restriction to functions of one real variable results in $\varepsilon(g) = \int_D (f(x) - g(x))^2\,dx$.

In case of a compact support $D = [a,b]$, the quantization problem is treated with a fixed boundary for the step functions, i.e., $a = x_1$ and $b = x_n$. In case of $D = \mathbb{R}$, the problem must be a free boundary problem, and in case of $D = \mathbb{R}_{\ge}$, it is semifree with $0 = x_1$. Lower indices $_0$ will always denote local optimal choices or optimal choices under some particular constraints like fixed partitions while upper indices ^{0} denote overall optimal choices.

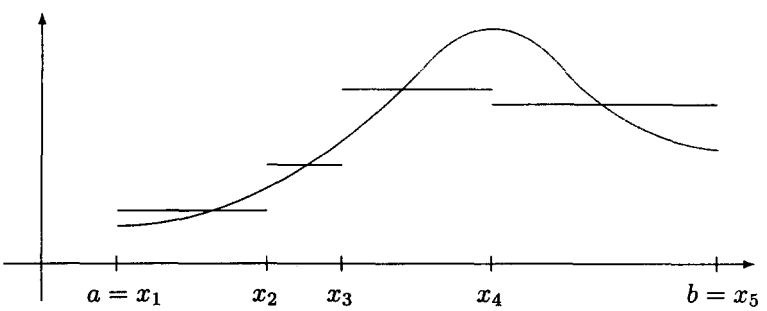

Figure 1. Function and its quantization.

LEMMA 1. *The minimum in the optimal quantization always exists.*

PROOF. Only levels y_i between infimum and supremum of f over D have to be considered. In case of a compact set $D = [a,b]$, these are finite with $e := \inf_{x\in[a,b]} f(x)$ and $h := \sup_{x\in[a,b]} f(x)$. The function $F(x_1,\ldots,x_n,y_1,\ldots,y_{n-1}) := \|f - \sum_{i=1}^n y_i 1_{(x_i,x_{i+1})}\|_p$ is continuous over the compact set $S_{\ge} \times [e,h]^{n-1}$, where $S_{\ge} = \{(x_1,\ldots x_n) \mid a = x_1 \le x_2 \le \cdots \le x_n = b\}$. The minimum of F over its domain thus, exists. If it does not satisfy inequalities between adjacent $x_i's$ strictly, then these $x_i's$ are slightly distorted with corresponding levels held constant. The minimal value of F is thus, not changed.

The argument is only sketched for $D \in \{\mathbb{R}_{\ge}, \mathbb{R}\}$. If a sequence of functions from S_n which approximates some f has a limit s in the sense of pointwise convergence, then $s \in S_n$. This eventually leads to $\inf_{s_n\in S_n} \|f - s_n\|_p = \|f - s\|_p$ for suitable $s \in S_n$. ∎

For $p = 2$ and $D = \mathbb{R}$, the objective under consideration is

$$\|f - s_n\|_2^2 = \int_{\mathbb{R}} (f(x) - s_n(x))^2\,dx = \int_{-\infty}^{x_1} f(x)^2\,dx + \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} (f(x) - y_i)^2\,dx + \int_{x_n}^{\infty} f(x)^2\,dx.$$

Reasonable choices for the levels y_i are generally not from the interval $[x_i, x_{i+1}]$, and hence, cannot be restricted to so-called regular quantizations, comp. [1, p. 135]. Usually the measure of fit in scalar quantization with respect to a probability density function f of a random variable X is taken to be the mean squared error

$$\|f - s_n\|_{\text{quant}} = \sum_{i=0}^{n} \int_{x_i}^{x_{i+1}} (x - y_i)^2 f(x)\, dx = E(X - s_n(X))^2,$$

where formally $x_0 = -\infty$ and $x_{n+1} = \infty$ are added to the set of break points of s_n. For a clear comparison of this approach to the one discussed below in the case $p = 2$, the additional levels have to be chosen to be zero: $y_0 = y_n = 0$. Both problems are related via integration by substitution.

LEMMA 2. *Let f be increasing and differentiable on $[x_i, x_{i+1}]$. Then*

$$\int_{x_i}^{x_{i+1}} (f(x) - y_i)^2\, dx = \int_{f(x_i)}^{f(x_{i+1})} (u - y_i)^2 \frac{1}{f'(f^{-1}(u))}\, du = \int_{f(x_i)}^{f(x_{i+1})} (u - y_i)^2 \left(f^{-1}\right)'(u)\, du.$$

PROOF. The equations result from substituting $f(x) = u$ in the first integral. ∎

Clearly, for a function f increasing on $[x_i, x_{i+1}]$ the level y_i should be chosen such that $f(x_i) \le y_i \le f(x_{i+1})$. Under suitable symmetry conditions this may lead to a regular quantization of $\mathbb{R}$ given by $\{f(x_i) \mid i = 1, \ldots, n\}$ instead of $\{x_i \mid i = 1, \ldots, n\}$. We do not try to adapt standard quantization procedures but rather develop direct ones. The main reason is that vector quantization methods hardly are globally convergent.

A probabilistic interpretation of the quantization problem is more complicated than that of vector quantization as mass preservation for quantization can only be guaranteed in special cases.

LEMMA 3. OPTIMAL QUANTIZATION IS MASS PRESERVING. *Let $p = 2$ and let $a = x_1 < \cdots < x_n = b$ be a partition of $D = [a, b]$. Any quantization $s_{n,0}$ which is optimal with respect to the given partition has the same mass as f, i.e., $\int_a^b f(x)\, dx = \int_a^b s_{n,0}(x)\, dx$.*

PROOF. Indicator functions of disjoint sets are orthogonal with respect to the inner product $(\cdot, \cdot)$ which induces the two-norm, where

$$(g, h) = \int_D g(x)h(x)\, dx.$$

Indicator functions $1_{(x_i, x_{i+1})}$ stemming from the fixed partition $x_1, \ldots, x_n$ span the linear space containing $s_{n,0}$. Hence, a standard argument for the error of a best approximation in terms of orthogonal expansions results in

$$\left(f - s_{n,0}, 1_{(x_i, x_{i+1})}\right) = 0, \qquad i = 1, \ldots, n-1.$$

Thus,

$$0 = \left(f - s_{n,0}, \sum_{i=1}^{n-1} 1_{(x_i, x_{i+1})}\right) = \left(f - s_{n,0}, 1_{(a,b)}\right) = \int_a^b f(x) - s_{n,0}(x)\, dx,$$

which completes the argument. ∎

If $p = 2$ and if f is a probability density function over $[a, b]$, then the optimal quantization s_n^0 also is a density and quantization of f amounts to approximating the given distribution by a finite discrete distribution over $[a, b]$. Quantization can thus be considered as constructing a histogram with an adaptive mesh for absolutely continuous distributions, comp. [13, Chapter 3]. Unlike in nonparametric statistics, the present construction and the measure of fit are not based on random samples but the construction of a sample is included.

Quantization in the sense of vector quantization is unbiased, i.e., for the optimal quantization s_n with respect to $\|\cdot\|_{\text{quant}}$ of density f belonging to a random variable X holds $EX = Es_n(X)$, [1, Formula (6.2.12), p. 180]. Optimal quantizations with respect to the p-norm are generally biased, even for $p = 2$.

To avoid trivial complications function f is assumed to be nonconstant in the sequel.

REMARK 1. For $1 \le p < \infty$ adjacent levels of an optimal quantization are distinct, i.e., $y^0_{i-1} \neq y^0_i$, $i = 2, \ldots, n-1$.

3. QUANTIZATION WITH PARTIALLY FIXED COMPONENTS

3.1. Fixed Partition

For a fixed partition the optimal levels can be characterized implicitly for general f in the case $p = 1$. The characterization is based on level (niveau) sets:

$$\begin{aligned}
N_=(y) &:= \{x \mid x \in [x_i, x_{i+1}] \text{ and } f(x) = y\},\\
N_>(y) &:= \{x \mid x \in [x_i, x_{i+1}] \text{ and } f(x) > y\},\\
N_<(y) &:= \{x \mid x \in [x_i, x_{i+1}] \text{ and } f(x) < y\},\\
N_\le(y) &:= N_<(y) \cup N_=(y),\\
N_\ge(y) &:= N_>(y) \cup N_=(y).
\end{aligned}$$

The level sets are not indexed by i for the sake of simplicity and they are measurable with respect to the Lebesgue measure λ as f is continuous.

THEOREM 1. DOMAIN BALANCING. *Let $p = 1$ and let $y_{i,0}$ denote the optimal quantization level for fixed $[x_i, x_{i+1}]$.*

1. *Assume $\lambda(N_=(y)) = 0$ for all $y \in \mathbb{R}$. Then*

$$\lambda\left(N_>(y_{i_0})\right) = \lambda\left(N_<(y_{i_0})\right) = \frac{x_{i+1} - x_i}{2}.$$

2. *$\lambda(N_=(y))$ arbitrary. Then $y_{i,0}$ equals the value*

$$y_{i,0} = \sup\left\{y \mid \lambda\left(N_<(y)\right) \le \frac{x_{i+1} - x_i}{2}\right\} = \inf\left\{y \mid \lambda\left(N_>(y)\right) \le \frac{x_{i+1} - x_i}{2}\right\}.$$

PROOF. The argument goes along variation of candidate levels.

PART 1. The effect of varying level y is captured by function H defined as

$$\begin{aligned}
H(y) &:= \int_{x_i}^{x_{i+1}} |f(x) - y|\, dx\\
&= \int_{N_>(y)} f(x)\, dx - \int_{N_<(y)} f(x)\, dx + y\left[\lambda\left(N_<(y)\right) - \lambda\left(N_>(y)\right)\right].
\end{aligned}$$

Without loss of generality it is supposed that $\lambda(N_<(y)) < \lambda(N_>(y))$. Then $\lambda(N_<(y)) + \lambda(N_>(y)) = x_{i+1} - x_i$. This implies $\lambda(N_<(y)) < (x_{i+1} - x_i)/2 < \lambda(N_>(y))$ as $\lambda(N_=(y)) = 0$. Select some $y' > y$ such that (also) $\lambda(N_<(y')) < \lambda(N_>(y'))$. It will be shown that $H(y') < H(y)$ demonstrating that y' is a better choice than y. The only choice for the level which cannot be improved is hence, $y_{i,0}$ giving equal Lebesgue measure to the upper and lower level sets. The difference of lower level sets is denoted by

$$N(y', y) := \{x \mid x \in [x_i, x_{i+1}] \text{ and } y \le f(x) < y'\} = N_<(y') - N_<(y).$$

Then $N_<(y') = N_<(y) \cup N(y', y)$ and $N_>(y) = N_>(y') \cup N(y', y)$. This results in

$$\begin{aligned}
H(y') < H(y) &\iff y' \left[\lambda(N_<(y')) - \lambda(N_>(y'))\right] - y \left[\lambda(N_<(y)) - \lambda(N_>(y))\right] \\
&< \int_{N_>(y)} f(x)\,dx - \int_{N_>(y')} f(x)\,dx - \left[\int_{N_<(y)} f(x)\,dx - \int_{N_<(y')} f(x)\,dx\right] \\
&\iff y' \left[\lambda(N_<(y')) - \lambda(N_>(y'))\right] \\
&\quad - y \left[\lambda(N_<(y')) - \lambda(N(y', y)) - \lambda(N_>(y')) - \lambda(N(y', y))\right] \\
&\quad < 2 \int_{N(y',y)} f(x)\,dx \\
&\iff (y' - y) \left[\lambda(N_<(y')) - \lambda(N_>(y'))\right] + 2y \cdot \lambda(N(y', y)) < 2 \int_{N(y',y)} f(x)\,dx \\
&\iff \underbrace{(y' - y)}_{>0} \cdot \underbrace{\left[\lambda(N_<(y')) - \lambda(N_>(y'))\right]}_{<0} < 2 \int_{N(y',y)} \underbrace{f(x) - y}_{\geq 0}\,dx,
\end{aligned}$$

where the last inequality is obvious.

PART 2. For

$$y_- := \sup \underbrace{\left\{ y \mid \lambda(N_\leq(y)) \leq \frac{x_{i+1} - x_i}{2} \right\}}_{=:Y_-}$$

and

$$y_+ := \inf \underbrace{\left\{ y \mid \lambda(N_\geq(y)) \leq \frac{x_{i+1} - x_i}{2} \right\}}_{=:Y_+}$$

the equality $y_- = y_+$ is demonstrated first. Therefore, a monotone increasing function $\varphi_\leq$ and a monotone decreasing function $\varphi_\geq$ are defined by

$$\begin{aligned}
\varphi_\leq(y) &:= \lambda(N_\leq(y)), \\
\varphi_\geq(y) &:= \lambda(N_\geq(y)).
\end{aligned}$$

Continuity of f implies $\varphi_\leq$ being strictly increasing, $\varphi_\geq$ being strictly decreasing on the range of f, and it implies lower bounds at the critical values y_- and y_+, respectively,

$$\varphi_\leq(y_-) = \lambda(N_\leq(y_-)) \geq \frac{x_{i+1} - x_i}{2} \qquad \text{and} \qquad \varphi_\geq(y_+) = \lambda(N_\geq(y_+)) \geq \frac{x_{i+1} - x_i}{2},$$

where both inequalities may be strict.

Assume $y_- < y_+$. Strict monotonicity of $\varphi_\leq$ and $\varphi_\geq$ imply for all w with $y_- < w < y_+$:

$$\varphi_\leq(w) > \frac{x_{i+1} - x_i}{2} \qquad \text{and} \qquad \varphi_\geq(w) > \frac{x_{i+1} - x_i}{2}.$$

Hence, $\lambda(N_=(w)) > 0$. This contradicts the functions $\varphi_\leq$ and $\varphi_\geq$ having at most a countable number of break points (since the functions are monotone) and these discontinuities being characterized by $\lambda(N_=(w)) > 0$.

Assume $y_- > y_+$. For all w with $y_+ < w < y_-$ there exist

1. w' with $w < w' < y_-$. Thus, strict monotonicity of $\varphi_\leq$ and $w' \in Y_-$ imply

$$\lambda(N_\leq(w)) < \lambda(N_\leq(w')) \leq \frac{x_{i+1} - x_i}{2}.$$

2. w'' with $y_+ < w'' < w'$. Thus, strict monotonicity of $\varphi_\geq$ and $w'' \in Y_+$ imply

$$\lambda(N_\geq(w)) < \lambda(N_\geq(w'')) \leq \frac{x_{i+1} - x_i}{2}.$$

Hence, $\lambda(N_\leq(w)) + \lambda(N_\geq(w)) < x_{i+1} - x_i \leq \lambda(N_\leq(w)) + \lambda(N_\geq(w))$, a contradiction.

$y_{i,0} = y_- = y_+$ can now be shown to satisfy the stated optimality by essentially the same swap argument as in Part 1. This argument requires the inequalities

$$\lambda(N_<(y)) < \lambda(N_>(y)), \quad \forall y < y_{i,0} \qquad \text{and} \qquad \lambda(N_>(y)) > \lambda(N_<(y)), \quad \forall y > y_{i,0}.$$

Without loss of generality we focus on the first inequality. Let $y < y_- = y_+$. Choose w with $y < w < y_-$.

1. Strict monotonicity of $\varphi_\leq$ and $w < y_-$ results in

$$\lambda(N_<(y)) \leq \lambda(N_\leq(y)) < \lambda(N_\leq(w)) \leq \frac{x_{i+1} - x_i}{2}.$$

2. Strict monotonicity of $\varphi_\geq$ and $w < y_+$ results in

$$\lambda(N_>(y)) \geq \lambda(N_\geq(y)) > \lambda(N_\geq(w)) > \frac{x_{i+1} - x_i}{2}.$$

This yields the desired inequality $\lambda(N_<(y)) < \lambda(N_>(y))$ completing the argument. ∎

The compact support $D = [a, b]$ is of special interest and will be considered from now on. For fixed partitions the optimal levels of the quantization can be stated explicitly in the most prominent cases $p = 2$ and $p = \infty$ and in the case $p = 1$ they can be given explicitly for monotone functions.

LEMMA 4. *For fixed x_i and x_{i+1} the induced optimal choice $y_{i,0}$ for level y_i is unique except for $p = \infty$ where $\|f - s_{n,0}\|_\infty$ may be attained on another interval than $[x_i, x_{i+1}]$.*

1. *Parameter $p = \infty$ results in*

$$y_{i,0} = \frac{1}{2}\left(\max_{x\in(x_i,x_{i+1})} f(x) + \min_{x\in(x_i,x_{i+1})} f(x)\right),$$

$$\max_{x\in(x_i,x_{i+1})} |f(x) - y_{i,0}| = \frac{1}{2}\left(\max_{x\in(x_i,x_{i+1})} f(x) - \min_{x\in(x_i,x_{i+1})} f(x)\right).$$

2. *Parameter $p = 2$ results in*

$$y_{i,0} = \frac{1}{x_{i+1} - x_i}\int_{x_i}^{x_{i+1}} f(x)\,dx = \frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i},$$

where F is the distribution function F of f; $F(x) = \int_a^x f(u)\,du$ for $x \in [a, b]$.

3. *Parameter $p = 1$ and f monotone and invertible or monotone and continuous on $[x_i, x_{i+1}]$ result in*

$$y_{i,0} = f\left(\frac{x_i + x_{i+1}}{2}\right).$$

PROOF.

PART 1. Let $f(x_{\max}) := \max_{x\in[x_i,x_{i+1}]} f(x)$, $f(x_{\min}) := \min_{x\in[x_i,x_{i+1}]} f(x)$, and $y_{i,0} := 1/2\,(f(x_{\max}) + f(x_{\min}))$. For any $y < y_{i,0}$ then holds

$$\max_{x\in[x_i,x_{i+1}]} |f(x) - y| = f(x_{\max}) - y > f(x_{\max}) - y_{i,0} = \frac{1}{2} f(x_{\max}) + \frac{1}{2} f(x_{\min}).$$

Similarly, for any $y > y_{i,0}$

$$\max_{x\in[x_i,x_{i+1}]} |f(x) - y| = y - f(x_{\min}) > y_{i,0} - f(x_{\min}) = \frac{1}{2} f(x_{\max}) + \frac{1}{2} f(x_{\min}).$$

Thus, $\min_y \max_{x\in[x_i,x_{i+1}]} |f(x) - y| = \max_{x\in[x_i,x_{i+1}]} |f(x) - y_{i,0}|$.

PART 2. Follows from the first orthogonality equation in the proof of Lemma 3 as

$$0 = \left(f - s_{n,0}, 1_{(x_i,x_{i+1})}\right) = \int_{x_i}^{x_{i+1}} f(x) - s_{n,0}(x)\,dx = \int_{x_i}^{x_{i+1}} f(x) - y_{i,0}\,dx$$
$$\Longrightarrow y_{i,0} = \frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} f(x)\,dx.$$

PART 3. Consequence of Theorem 1 as in the case of function f being monotone and invertible $\lambda(N_=(y)) = 0$, $\forall y$. Hence, $\lambda(N_<(y_{i0})) = \lambda(N_>(y_{i0})) \Longleftrightarrow \lambda([x_i, f^{-1}(y_{i0}))) = \lambda((f^{-1}(y_{i0}), x_{i+1}])$ $\Longleftrightarrow f^{-1}(y_{i0}) - x_i = x_{i+1} - f^{-1}(y_{i0}) \Longleftrightarrow f^{-1}(y_{i0}) = (x_i + x_{i+1})/2$.

For monotone increasing and continuous f the set theoretic inverse $f^{-1}_{\text{set}}(y) = \inf\{x \mid f(x) = y\}$ exists and $N_<(y) = \{x \mid f(x) < y\} = \{x \mid x < f^{-1}_{\text{set}}(y)\} = [x_i, f^{-1}_{\text{set}}(y))$ as well as $N_>(y) = \{x \mid f(x) > y\} = (f^{-1}_{\text{set}}(y), x_{i+1}]$. Hence, Theorem 1 results in

$$y_{i,0} = \sup\left\{y \mid \lambda\left(N_<(y)\right) \le \frac{x_{i+1} - x_i}{2}\right\} = \sup\left\{y \mid f^{-1}_{\text{set}}(y) - x_i \le \frac{x_{i+1} - x_i}{2}\right\}$$
$$= \sup\left\{y \mid f^{-1}_{\text{set}}(y) \le \frac{x_{i+1} + x_i}{2}\right\}$$

and

$$y_{i,0} = \inf\left\{y \mid \lambda\left(N_>(y)\right) \le \frac{x_{i+1} - x_i}{2}\right\} = \inf\left\{y \mid x_{i+1} - f^{-1}_{\text{set}}(y) \le \frac{x_{i+1} - x_i}{2}\right\}$$
$$= \inf\left\{y \mid f^{-1}_{\text{set}}(y) \ge \frac{x_{i+1} + x_i}{2}\right\}.$$

Continuity of f then implies $f((x_{i+1} + x_i)/2) = y_{i,0}$. As f may be constant in level $y_{i,0}$ the arithmetic mean of the interval boundaries may not be the only solution to this level equation in which case $f^{-1}_{\text{set}}(y) \le (x_{i+1} + x_i)/2$. ∎

The idea of balancing applies throughout all finite values of p as function f contains enough variability, especially as it is not constant over some segment.

LEMMA 5. VALUE BALANCING. *Let $\lambda(N_=(y)) = 0$, $\forall y \in \mathbb{R}$. The optimal quantization level for a fixed partition and integer value p with $1 \le p < \infty$ is then given by*

$$\int_{N_<(y)} (y - f(x))^{p-1}\,dx = \int_{N_>(y)} (f(x) - y)^{p-1}\,dx.$$

PROOF. Function H defined by $H(y) := \int_{x_i}^{x_{i+1}} |f(x) - y|^p\,dx$ is differentiable as the integrand is differentiable almost everywhere in $[x_i, x_{i+1}]$ and

$$\frac{d}{dy} H(y) = \int_{x_i}^{x_{i+1}} \frac{d}{dy} |f(x) - y|^p\,dx$$

with

$$\frac{d}{dy} |f(x) - y|^p = \begin{cases} -p(f(x) - y)^{p-1}, & \text{if } y < f(x), \\ p(y - f(x))^{p-1}, & \text{if } y > f(x). \end{cases}$$

Hence,

$$H'(y) = 0 \Longleftrightarrow \int_{N_>(y)} -p(f(x) - y)^{p-1}\,dx + \int_{N_<(y)} p(y - f(x))^{p-1}\,dx = 0$$
$$\Longleftrightarrow \int_{N_<(y)} (y - f(x))^{p-1}\,dx = \int_{N_>(y)} (f(x) - y)^{p-1}\,dx.$$ ∎

The formula of Lemma 5 results in the expression of Theorem 1.1 for $p = 1$ with $(f(x) - y)^0 = 1$, where $0^0 = 1$. In the Hilbert case $p = 2$ the formula specializes to an implicit version of Lemma 4.2 for increasing and invertible f:

$$\int_{x_i}^{f^{-1}(y)} y - f(x)\,dx = \int_{f^{-1}(y)}^{x_{i+1}} f(x) - y\,dx.$$

3.2. Fixed Levels

The levels of the step functions are tentatively considered to be fixed. For $p = \infty$ there is no Chebychev alternation theorem. As a consequence, the idea behind the second algorithm by Remes (see, e.g., [14, p. 97]) on optimal placement of break points for best approximating polynomials fails to be applicable here.

For monotone functions and fixed levels the quantization problem is solvable uniformly in p. This uniformity is a break in symmetry between the cases of fixed levels and fixed partitions, since for fixed partitions the optimal levels depend on p even for monotone functions f.

LEMMA 6. *Let $1 \le p \le \infty$, let f be increasing and differentiable, and let all levels $y_1, \dots, y_{n-1}$ be fixed with $f(a) < y_1 < \cdots < y_{n-1} < f(b)$. An optimal partition $a = x_{1,0} < x_{2,0} < \cdots < x_{n-1,0} < x_{n,0} = b$ is then given implicitly by*

$$f(x_{i,0}) = \frac{y_{i-1} + y_i}{2}, \qquad i = 2, \dots, n-1.$$

PROOF. In case $1 \le p < \infty$ the p^{th} power of the objective function is $G(x_2, \dots, x_{n-1}) = \|f - s_n\|_p^p = \sum_{j=1}^{n-1} \int_{x_j}^{x_{j+1}} |f(x) - y_j|^p \, dx$. For $i = 2, \dots, n-1$ the derivatives thus, yield

$$\begin{aligned}
\frac{\partial}{\partial x_i} G(x_2, \dots, x_{n-1}) &= \frac{\partial}{\partial x_i} \left[\int_{x_{i-1}}^{x_i} |f(x) - y_{i-1}|^p \, dx + \int_{x_i}^{x_{i+1}} |f(x) - y_i|^p \, dx \right] \\
&= |f(x_i) - y_{i-1}|^p - |f(x_i) - y_i|^p \stackrel{!}{=} 0 \\
&\iff f(x_i) - y_{i-1} = y_i - f(x_i) \\
&\iff f(x_i) = \frac{y_{i-1} + y_i}{2}.
\end{aligned}$$

The Hessian of G is a diagonal matrix with the nonnegative entries for all critical points of G

$$\frac{\partial^2}{\partial^2 x_i} G(x_2, \dots, x_{n-1}) = p f'(x_i) \left[(f(x_i) - y_{i-1})^{p-1} + (y_i - f(x_i))^{p-1} \right] \ge 0, \qquad i = 2, \dots, n-1.$$

Thus, a critical point of G is a local minimum.

In case $p = \infty$ the minimization can also be reduced to adjacent levels:

$$\begin{aligned}
\min_{x_2, \dots, x_{n-1}} \|f - s_n\|_\infty = \min_{x_2, \dots, x_{n-1}} \max \Big\{ & y_1 - f(a), \\
& \underbrace{f(x_2) - y_1, y_2 - f(x_2)}, \dots, \\
& \underbrace{f(x_{n-1}) - y_{n-2}, y_{n-1} - f(x_{n-1})}, \\
& f(b) - y_{n-1} \Big\} \\
= \max \Big\{ & y_1 - f(a), \max_{i=2,\dots,n-1} \min_{x_i} \max\{f(x_i) - y_{i-1}, y_i - f(x_i)\}, \\
& f(b) - y_{n-1} \Big\}.
\end{aligned}$$

Monotonicity of f implies that each term $\max\{f(x_i) - y_{i-1}, y_i - f(x_i)\}$ is minimized over x_i for $f(x_i) - y_{i-1} = y_i - f(x_i) \iff f(x_i) = (y_{i-1} + y_i)/2$. ∎

Solutions $x_{i,0}$ in the previous lemma need not be unique for $i = 2, \dots, n-1$. The monotonicity condition is removed next. Uniformity in p of all candidates for an optimal partition are preserved which is shown essentially by the same argument as for the increasing case.

LEMMA 7. *Let $1 \le p \le \infty$ and let f be differentiable on D. Suppose successive levels are distinct, i.e., $y_{i-1} \neq y_i$, $i = 2, \dots, n-1$. The unknown break points $x_{i,0}$ of $s_{n,0}$ satisfy the condition*

$$[x_{i,0} \in \{a, b\}] \text{ or } \left[f(x_{i,0}) = \frac{y_{i-1} + y_i}{2} \text{ and } \operatorname{sgn} f'(x_{i,0}) = \operatorname{sgn}(y_i - y_{i-1}) \right], \qquad i = 2, \dots, n-1.$$

The case $x_{i,0} \in \{a, b\}$ accounts for less than $n-2$ solutions of the level and sign conditions. The sign condition on f' states that f and $s_{n,0}$ are both either increasing or decreasing at $x_{i,0}$. Sufficiency for global optimality subject to fixed levels is obtained by choosing from the local optima.

LEMMA 8. *Let $n = 3$ and let several locations $x_{2,1} < \cdots < x_{2,r}$ satisfy the local optimality criteria of Lemma 7. The global minimum $x_{2,0}$ among them is given*

1. *in case $p = 2$ by*

$$\min_{1 \le j \le r} \operatorname{sgn}(y_2 - y_1) \cdot \int_{x_{2,1}}^{x_{2,j}} f(x) - \frac{y_1 + y_2}{2}\, dx = \operatorname{sgn}(y_2 - y_1) \cdot \int_{x_{2,1}}^{x_{2,0}} f(x) - \frac{y_1 + y_2}{2}\, dx;$$

2. *in case $1 \le p < \infty$ $(p \neq 2)$ by*

$$\min_{1 \le j \le r} \int_a^{x_{2,j}} |f(x) - y_1|^p\, dx + \int_{x_{2,j}}^b |f(x) - y_2|^p\, dx = \int_a^{x_{2,0}} |f(x) - y_1|^p\, dx + \int_{x_{2,0}}^b |f(x) - y_2|^p\, dx;$$

3. *in case $p = \infty$ by*

$$x_{2,0} \in \{x_{2,1}, \dots, x_{2,r}\} \text{ arbitrary.}$$

Simplifying the expression of Lemma 8.2 is not obvious for arbitrary p. For $p = 1$ and $y_1 < y_2$ the expression is equivalent to $x_{2,0}$ solving the finite minimization problem

$$\min_{x_{2,j}: 1 \le j \le r} 2 \int_{\{x \mid x \in [x_{2,1}, x_{2,j}] \text{ and } y_1 \le f(x) < y_2\}} f(x) - \frac{y_1 + y_2}{2}\, dx$$
$$+ (y_2 - y_1) \cdot [\lambda(\{x \mid x \in [x_{2,1}, x_{2,j}] \text{ and } f(x) > y_2\}) - \lambda(\{x \mid x \in [x_{2,1}, x_{2,j}] \text{ and } f(x) < y_1\})].$$

Even the outer candidate locations $x_{2,1}$ and $x_{2,r}$ can the best choices.

4. QUANTIZATION WITH VARIABLE PARTITION AND VARIABLE LEVELS

Computations of optimal levels given a partition (Lemma 4) combined with computing a partition for given levels (Lemma 7) results for $p = 1$, 2, or ∞ in an iterative alternating variable procedure for quantization on $D = [a, b]$, where the boundary is fixed at $x_1 = a$ and $x_n = b$. The Hilbert case $p = 2$ is focussed on from now which results in the subsequent generic procedure.

GenP

1. Initialization. An arbitrary partition $a = x_1^{(1)} \le x_2^{(1)} \le \cdots \le x_{n-1}^{(1)} \le x_n^{(1)} = b$ is selected. Set $k \leftarrow 1$.
2. Iteration. Repetition until some stopping criterion is met:
   (a) Computation of levels $y_i^{(k)}$ by

$$y_i^{(k)} = y\left(x_i^{(k)}, x_{i+1}^{(k)}\right) := \frac{1}{x_{i+1}^{(k)} - x_i^{(k)}} \int_{x_i^{(k)}}^{x_{i+1}^{(k)}} f(x)\, dx = \frac{F\left(x_{i+1}^{(k)}\right) - F\left(x_i^{(k)}\right)}{x_{i+1}^{(k)} - x_i^{(k)}}, \qquad i = 1, \dots, n-1$$

   with $F(x) = \int_a^x f(u)\, du$.

(b) Computation of a new partition with $x_i^{(k+1)}$ such that

$$f\left(x_i^{(k+1)}\right) = \frac{y_{i-1}^{(k)} + y_i^{(k)}}{2} \quad \text{and}$$
$$\begin{cases} \operatorname{sgn} f'\left(x_i^{(k+1)}\right) = \operatorname{sgn}\left(y_i^{(k)} - y_{i-1}^{(k)}\right) \neq 0, & \text{if such } x_i^{(k+1)} \text{ exists,} \\ \text{different from } x_i^{(k)}, a, \text{ and } b, & \text{else,} \end{cases} \qquad i = 2, \ldots, n-1,$$

(c) $k \leftarrow k+1$.

Level function $y(x_i^{(k)}, x_{i+1}^{(k)})$ is set equal to the value $f(x_i^{(k)})$ in case $x_{i+1}^{(k)} = x_i^{(k)}$. If f is not differentiable in $x_i^{(k+1)}$, then $\operatorname{sgn} f'(x_i^{(k+1)})$ is replaced by $+1$ or -1 for f being increasing or decreasing in $x_i^{(k+1)}$, respectively. The level function $y(x_i, x_{i+1})$ depicts the optimal quantization level for interval $[x_i, x_{i+1}]$.

The computations in Step 2(a) are unique while the selections in Step 2(b) are generally not. In case f is not monotone, a partition of $[a, b]$ may even lead to adjacent levels of equal value. An example is $f(x) = \sin x$ over $[a, b] = [0, 4\pi]$ with $n = 3$ and $x_2^{(1)} = 2\pi$ which leads to $y_1 = y_2 = 1/2\pi \cdot \int_0^{2\pi} \sin x \, dx = 0 = \sin x_2^{(2)}$ so that $x_2^{(2)} \in \{0, \pi, 2\pi, 3\pi, 4\pi\}$. **GenP** chooses $x_2^{(2)}$ arbitrarily from $\{\pi, 3\pi\}$.

The stopping criterion in **GenP** can be based on successive differences. The procedure may for example terminate if $\max_{i=2,\ldots,n-1} \{|x_i^{(k)} - x_i^{(k+1)}|\}$ does not exceed a given threshold. This carries over to subsequent specializations of **GenP**.

For a partition $a = x_1 < x_2 < \cdots < x_{n-1} < x_n = b$ **GenP** computes in each iteration the quantization with optimal levels resulting in the objective

$$\mu_f(x_2, \ldots, x_{n-1}) := \mu(x_2, \ldots, x_{n-1}) := \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} (f(x) - y(x_i, x_{i+1}))^2 \, dx$$

with (partial) derivatives for $i = 2, \ldots, n-1$

$$\frac{\partial}{\partial x_i} \mu(x_2, \ldots, x_{n-1}) = (y(x_i, x_{i+1}) - y(x_{i-1}, x_i)) \cdot (2f(x_i) - y(x_i, x_{i+1}) - y(x_{i-1}, x_i)),$$
$$\frac{\partial}{\partial x_i} y(x_i, x_{i+1}) = \frac{-f(x_i)}{x_{i+1} - x_i} + \frac{y(x_i, x_{i+1})}{x_{i+1} - x_i}, \qquad \text{and}$$
$$\frac{\partial}{\partial x_i} y(x_{i-1}, x_i) = \frac{f(x_i)}{x_i - x_{i-1}} - \frac{y(x_{i-1}, x_i)}{x_i - x_{i-1}}.$$

The objective $\mu(x_2, \ldots, x_{n-1})$ is a continuous function defined over the simplicial set $\Delta := \Delta_n := \Delta([a, b]) := \{(w_2, \ldots, w_{n-1})^\top \mid a \leq w_2 \leq \cdots \leq w_{n-1} \leq b\}$, where $\Delta([a, b]) = [a, b]$ in the two-level case $n = 3$.

LEMMA 9. *The sequence of objectives* $(\mu(x_2^{(k)}, \ldots, x_{n-1}^{(k)}))_{k=1}^{\infty}$ *for partitions generated by* **GenP** *is decreasing for an arbitrary initial partition.*

PROOF. Straightforward from the construction of **GenP** and Lemmata 4 and 7. ∎

As a consequence, whenever for some function f a point $(x_2, \ldots, x_{n-1})$ will be generated by an associated fixed-point function g_f as below, this point is not periodic in the sense of iterated function systems except in the trivial case where it is a fixed point. Thus, from the viewpoint of dynamical systems quantization is simple—for example no bifurcations can occur whatever a parametrized family $(f_\lambda)_{\lambda \in \Lambda}$ should look like. Moreover, considerations are restricted to functions f with finite number of fixed points of g_f so that chaotic behaviour can be excluded.

4.1. Monotone Functions

The quantization problem is easiest to investigate for monotone f. Without loss of generality f is assumed to be nonnegative. Monotone functions are chosen to be increasing and integrating to unity over D if not stated otherwise. A decreasing f can be substituted by an increasing f^* with $f^*(a+b-x)$. The level function $y(v,w)$ is then increasing in both arguments. The increasing function f is furthermore assumed to be not constant over any interval. Optimal quantizations satisfy an interlace condition if the number of levels varies by one. This will be shown by a gradient projection argument. For the sake of simplicity the argument is stated separately for the two-level case.

LEMMA 10. INTERLACING OF LOCAL MINIMA FOR MONOTONE f. *For each local minimum $x_{2,0}(3)$ of $\mu_f(x_2)$ there is a local minimum $(x_{2,0}(4), x_{3,0}(4))$ of $\mu_f(x_2,x_3)$ for f monotone such that the minima are interlaced*

$$x_{2,0}(4) < x_{2,0}(3) < x_{3,0}(4)$$

and for each local minimum of $\mu_f(x_2,x_3)$ there is a local minimum of $\mu_f(x_2)$ such that the interlace condition is satisfied.

PROOF. Let $x_{2,0}(3)$ be a local minimum of $\mu_f(x_2)$. The set $\Gamma_4 := \Gamma_4(x_{2,0}(3)) := \{(x_2,x_3)^\top \mid a \le x_2 \le x_{2,0}(3) \le x_3 \le b\}$ is considered. The value $\min_{(x_2,x_3)\in\Gamma_4} \mu_f(x_2,x_3)$ will be shown to be attained in $\mathrm{int}(\Gamma_4)$. As a neighbourhood in Γ_4 is also one in Δ_4, local minimality extends from Γ_4 to Δ_4 which will complete the argument.

For $(x_2,x_3) \in \partial(\Gamma_4)$ with $x_2 = a$ or $x_3 = b$ a slight transition from $\partial(\Gamma_4)$ into $\mathrm{int}(\Gamma_4)$ obviously results in a decrease of $\mu_f(x_2,x_3)$. Thus, the two cases with $x_2 = x_{2,0}(3)$ and $x_3 = x_{2,0}(3)$ remain to be analyzed where the first case is considered only without loss of generality, comp. Figure 2.

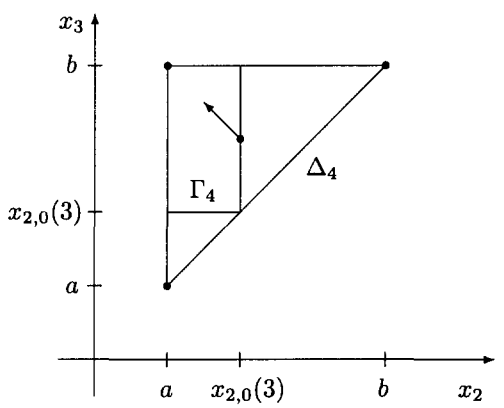

Figure 2. Negative gradient pointing into $\mathrm{int}(\Gamma_4)$.

For each $(x_{2,0}(3), x_3') \in \Gamma_4$ there is $(x_2,x_3) \in \mathrm{int}(\Gamma_4)$ such that $\mu_f(x_2,x_3) < \mu_f(x_{2,0}(3),x_3')$. This is derived from the univariate objective $\mu_f(x_2)$ having a local minimum in $x_{2,0}(3)$, hence,

$$\left.\frac{d\mu_f(x_2)}{dx_2}\right|_{x_2=x_{2,0}(3)} = \underbrace{(y(x_{2,0}(3),b) - y(a,x_{2,0}(3)))}_{>0} \cdot (2f(x_{2,0}(3)) - y(a,x_{2,0}(3)) - y(x_{2,0}(3),b)) = 0,$$

and thus, $2f(x_{2,0}(3)) - y(a,x_{2,0}(3)) = y(x_{2,0}(3),b) > y(x_{2,0}(3),x_3)$ for $x_3 < b$. This results in

$$\left.\frac{\partial\mu_f(x_2,x_3)}{\partial x_2}\right|_{(x_{2,0}(3),x_3)} = \underbrace{(y(x_{2,0}(3),x_3) - y(a,x_{2,0}(3)))}_{>0} \cdot \underbrace{(2f(x_{2,0}(3)) - y(a,x_{2,0}(3)) - y(x_{2,0}(3),x_3))}_{>0} > 0.$$

Hence, the negative gradient of μ_f points from each $(x_{2,0}(3), x_3') \in \partial(\Gamma_4)$ with $x_{2,0}(3) < x_3' < b$ into $\mathrm{int}(\Gamma_4)$ resulting in the desired inequality. Also, the values of $\mu_f(x_2, x_3)$ at $(x_{2,0}(3), x_{2,0}(3))$ and $(x_{2,0}(3), b)$ are obviously decreased when slightly moving from these points along their connecting line which lies on $\partial(\Gamma_4)$.

The argument is reversed for each local minimum of $\mu_f(x_2, x_3)$ to interlace one of $\mu_f(x_2)$. ∎

Several local minima of $\mu_f(x_2)$ may be interlaced by the same local minimum of $\mu_f(x_2, x_3)$ which is the case in Example 2 below.

THEOREM 2. INTERLACING OF LOCAL MINIMA FOR MONOTONE f. *For each local minimum* $(x_{2,0}(n), \ldots, x_{n-1,0}(n))$ *of* $\mu_f(x_2, \ldots, x_{n-1})$ *for an increasing function* f *there is an interlacing local minimum* $(x_{2,0}(n+1), \ldots, x_{n,0}(n+1))$ *of* $\mu_f(x_2, \ldots, x_n)$, *i.e.,*

$$a < x_{2,0}(n+1) < x_{2,0}(n) < x_{3,0}(n+1) < x_{3,0}(n) < \cdots < x_{n-1,0}(n) < x_{n,0}(n+1) < b$$

and for each local minimum of $\mu_f(x_2, \ldots, x_n)$ *there is a local minimum of* $\mu_f(x_2, \ldots, x_{n-1})$ *such that the interlace condition is satisfied.*

PROOF. For the local minimum $(x_{2,0}(n), \ldots, x_{n-1,0}(n))$ of $\mu_f(x_2, \ldots, x_{n-1})$ the set $\Gamma_{n+1} := \Gamma_{n+1}(x_{2,0}(n), \ldots, x_{n-1,0}(n)) := \{(x_2, \ldots, x_n)^\top \mid a \le x_2 \le x_{2,0}(n) \le x_3 \le \cdots \le x_{n-1} \le x_{n-1,0}(n) \le x_n \le b\}$ is considered. $\Gamma_{n+1} \subseteq \Delta_{n+1}$. For each element $(x_2, \ldots, x_n)$ of Γ_{n+1} with at least one interlacing inequality satisfied as equality there is one element of Γ_{n+1} satisfying all interlacing conditions strictly and having a smaller quantization error. The possible equations $x_2 = a$ and $x_n = b$ can obviously be distorted to $a < x_2$ and $x_n < b$ to result in a decrease of $\mu_f(x_2, \ldots, x_n)$.

Let $i \in \{2, \ldots, n\}$ denote the smallest index such that x_i is at its upper bound in Γ_{n+1}.

CASE 1. $i = n$. Then $x_2 < x_{2,0}(n), \ldots, x_{n-1} < x_{n-1,0}(n), x_n = b$. Obviously, a decrease in $\mu_f(x_2, \ldots, x_n)$ results from decreasing x_n.

CASE 2. $3 \le i \le n-1$. Then $x_2 < x_{2,0}(n), \ldots, x_{i-1} < x_{i-1,0}(n), x_i = x_{i,0}(n)$. Hence,

$$\left.\frac{\partial \mu_f(x_2, \ldots, x_n)}{\partial x_i}\right|_{(x_2, \ldots, x_i = x_{i,0}(n), \ldots, x_n)} = \underbrace{(y(x_{i,0}(n), x_{i+1}) - y(x_{i-1}, x_{i,0}(n)))}_{>0} \cdot \underbrace{(2f(x_{i,0}(n)) - y(x_{i-1}, x_{i,0}(n)) - y(x_{i,0}(n), x_{i+1}))}_{>0} > 0,$$

because the point $(x_{2,0}(n), \ldots, x_{n-1,0}(n))$ being a critical point of $\mu_f(x_2, \ldots, x_{n-1})$ results in

$$2f(x_{i,0}(n) - y(x_{i-1,0}(n), x_{i,0}(n)) - y(x_{i,0}(n), x_{i+1,0}(n)) = 0$$

and the definition of $i \ge 3$ implies

$$\begin{aligned} x_{i-1} < x_{i-1,0}(n) &\Longrightarrow y(x_{i-1}, x_{i,0}(n)) < y(x_{i-1,0}(n), x_{i,0}(n)) \\ x_{i+1} \le x_{i+1,0}(n) &\Longrightarrow y(x_{i,0}(n), x_{i+1}) \le y(x_{i,0}(n), x_{i+1,0}(n)) \\ &\Longrightarrow 2f(x_{i,0}(n)) - y(x_{i-1}, x_{i,0}(n)) - y(x_{i,0}(n), x_{i+1}) > 0. \end{aligned}$$

The negative gradient of $\mu_f(x_2, \ldots, x_n)$ thus, tends to decrease x_i.

CASE 3. $i = 2$. The index $j \in \{i+1, \ldots, n\}$ is defined as the smallest index such that $x_{i+1}, \ldots, x_j$ are at their upper bounds in Γ_{n+1}. If this index does not exist, then $x_3 < x_{3,0}(n)$ and

$$\left.\frac{\partial \mu_f(x_2, \ldots, x_n)}{\partial x_2}\right|_{(x_{2,0}(n), x_3, \ldots, x_n)} = (y(x_{2,0}(n), x_3) - y(a, x_{2,0}(n))) \cdot (2f(x_{2,0}(n)) - y(a, x_{2,0}(n)) - y(x_{2,0}(n), x_3)).$$

The inequality $x_3 < x_{3,0}(n)$ results in $y(x_{2,0}(n), x_3) < y(x_{2,0}(n), x_{3,0}(n))$, and hence, in the inequality $2f(x_{2,0}(n)) - y(a, x_{2,0}(n)) - y(x_{2,0}(n), x_3) > 0$. The negative gradient tends to decrease x_2.

If there is an index j, then iterated applications of Cases 1 and 2 lead to a successive decrease of $x_j, x_{j-1}, \ldots, x_3$ and eventually to a decrease of x_2. Repeated application of alterations of $(x_2, \ldots, x_n)$ stated in all the cases leads to all inequalities from Γ_{n+1} to be strictly satisfied.

The argument can essentially be reversed for showing that a local minimum of $\mu_f(x_2, \ldots, x_n)$ interlaces one of $\mu_f(x_2, \ldots, x_{n-1})$. ∎

Monotonicity of f transforms the generic procedure **GenP** to the subsequent procedure **FP** which is refered to as fixed-point procedure for reasons to become obvious soon.

FP

1. Initialization. An arbitrary partition $a = x_1^{(1)} \leq x_2^{(1)} \leq \cdots \leq x_{n-1}^{(1)} \leq x_n^{(1)} = b$ is selected. Set $k \leftarrow 1$.
2. Iteration. Repetition until some stopping criterion is met:
   (a) Computation of levels $y_i^{(k)} = y(x_i^{(k)}, x_i^{(k+1)})$, $i = 1, \ldots, n-1$.
   (b) Computation of new partition $x_i^{(k+1)}$ by $f(x_i^{(k+1)}) = (y_{i-1}^{(k)} + y_i^{(k)})/2$, $i = 2, \ldots, n-1$.
   (c) $k \leftarrow k + 1$.

REMARK 2. *Function f being increasing, respectively, decreasing results in*

1. $y_1^{(k)} \leq \cdots \leq y_{n-1}^{(k)}$, *respectively,* $y_1^{(k)} \geq \cdots \geq y_{n-1}^{(k)}$, $\forall k$ *and*
2. $y_{i-1}^{(k)} \leq f(x_i^{(k)}) \leq y_i^{(k)}$, *respectively,* $y_{i-1}^{(k)} \geq f(x_i^{(k)}) \geq y_i^{(k)} i = 2, \ldots, n-1$, $\forall k$.

Each monotone and invertible function f is assigned a continuous fixed-point function g_f over the simplicial set Δ with values being defined for $n \geq 3$ by

$$g_f \begin{pmatrix} x_2 \\ \vdots \\ x_{n-1} \end{pmatrix} := \begin{pmatrix} f^{-1}\left(\frac{(y_1 + y_2)}{2}\right) \\ \vdots \\ f^{-1}\left(\frac{(y_{n-2} + y_{n-1})}{2}\right) \end{pmatrix} := \begin{pmatrix} f^{-1}\left(\frac{(y(a, x_2) + y(x_2, x_3))}{2}\right) \\ \vdots \\ f^{-1}\left(\frac{(y(x_{n-2}, x_{n-1}) + y(x_{n-1}, b))}{2}\right) \end{pmatrix}.$$

LEMMA 11. *Monotone functions admit the equivalence between fixed points and critical points*

$$g_f(x_2, \ldots, x_{n-1}) = (x_2, \ldots, x_{n-1})^\top \iff \operatorname{grad}(\mu_f(x_2, \ldots, x_{n-1})) = 0 \text{ over } \operatorname{int}(\Delta).$$

PROOF. Function f being increasing implies $y(x_{i-1}, x_i) < y(x_i, x_{i+1})$ for all triplets x_{i-1}, x_i, x_{i+1} with $x_{i-1} < x_{i+1}$, $i = 2, \ldots, n-1$. Hence,

$$\begin{aligned} & g_f(x_2, \ldots, x_{n-1}) = (x_2, \ldots, x_{n-1})^\top \\ & \quad \iff 2f(x_i) = y(x_{i-1}, x_i) + y(x_i, x_{i+1}), \qquad \forall i = 2, \ldots, n-1 \\ \iff & \frac{\partial}{\partial x_i} \mu_f(x_2, \ldots, x_{n-1}) = (y(x_i, x_{i+1}) - y(x_{i-1}, x_i)) \cdot (2f(x_i) - y(x_i, x_{i+1}) - y(x_{i-1}, x_i)) \\ & \quad = 0, \qquad \forall i = 2, \ldots, n-1. \end{aligned}$$

∎

Extrema of the objective are related to the attracting behaviour of fixed points. A fixed point is contracting, if there is a neighborhood U around it such that the sequence of successive approximations converges to the fixed point for each initial value in U. The set of all values from which the successive approximations converge towards a fixed point x is its domain of attraction $\operatorname{dom}(x)$. A fixed point which is not contracting is repelling. A special case of repellation is half-sided repellation and half-sided attraction.

LEMMA 12. *A contracting fixed point of the function g_f is a local minimum of μ_f.*

PROOF. Let x be a contracting fixed point of g_f. There is thus, a neighbourhood $U = U(x)$ such that $(x^{(k)})_{k=1}^{\infty}$ converges to x $\forall x^{(1)} \in U$. Then $\mu_f(x) \le \mu_f(z)$, $\forall z \in U$ because if $\exists x^{(1)} \in U$ with $\mu_f(x^{(1)}) < \mu_f(x)$ then $\mu_f(x) = \lim_{k\to\infty} \mu_f(x^{(k)}) \le \mu_f(x^{(1)}) < \mu_f(x)$, a contradiction. ∎

For later use we establish a Lipschitz bound of the quantization objective μ_f. As $\operatorname{grad}\mu_f$ is on the same "integration level" as f, the quantization objective μ_f has a Lipschitz bound even if f is not known to have one.

LEMMA 13. *Let f be increasing over $D = [a, b]$. Then*

1. $\|\operatorname{grad}\mu_f(x_2, \ldots, x_{n-1})\| \le (n-2)(f(b) - f(a))^2$, $\forall (x_2, \ldots, x_{n-1}) \in \Delta([a, b])$ *and*
2. *if f has Lipschitz bound L, then* $\|\operatorname{grad}\mu_f(x_2, \ldots, x_{n-1})\| \le (n-2)(L/2 \cdot (b-a))^2$, $\forall (x_2, \ldots, x_{n-1}) \in \Delta([a, b])$.

PROOF. The formula for partial derivatives of the quantization objective implies

$$\begin{aligned}\left|\frac{\partial\mu_f(x_2, \ldots, x_{n-1})}{\partial x_i}\right| &\le |y(x_i, x_{i+1}) - y(x_{i-1}, x_i)| \cdot |2f(x_i) - y(x_i, x_{i+1}) - y(x_{i-1}, x_i)| \\ &\le (y(x_i, x_{i+1}) - y(x_{i-1}, x_i))(f(x_i) - y(x_{i-1}, x_i) + y(x_i, x_{i+1}) - f(x_i)) \\ &= (y(x_i, x_{i+1}) - y(x_{i-1}, x_i))^2 \le (f(b) - f(a))^2.\end{aligned}$$

PART 1. The Hölder inequality results in

$$\|\operatorname{grad}\mu_f(x_2, \ldots, x_{n-1})\| \le \sum_{i=2}^{n-1} \left|\frac{\partial\mu_f(x_2, \ldots, x_{n-1})}{\partial x_i}\right| \le (n-2)(f(b) - f(a))^2.$$

PART 2. The Lipschitz bound for f implies

$$\begin{aligned}0 \le y(x_i, x_{i+1}) - y(x_{i-1}, x_i) &= \frac{\int_{x_i}^{x_{i+1}} f(u)\,du}{x_{i+1} - x_i} - \frac{\int_{x_{i-1}}^{x_i} f(u)\,du}{x_i - x_{i-1}} \\ &\le \frac{\int_{x_i}^{x_{i+1}} f(x_i) + L(u - x_i)\,du}{x_{i+1} - x_i} - \frac{\int_{x_{i-1}}^{x_i} f(x_i) - L(x_i - u)\,du}{x_i - x_{i-1}} \\ &= f(x_i) + \frac{L}{2}(x_{i+1} - x_i) - \left(f(x_i) - \frac{L}{2}(x_i - x_{i-1})\right) \\ &= \frac{L}{2}(x_{i+1} - x_{i-1}) \le \frac{L}{2}(b - a).\end{aligned}$$

∎

There appears to be no general improvement for the Lipschitz bounds from Lemma 13.1 and Lemma 13.2 on the complete interval $[a, b]$, since they become sharp even in the two-level case $n = 3$ for linear functions f at $x = a$ and $x = b$. Obvious improvements of the Lipschitz bound result from restrictions to subintervals. Increasing functions in the two-level case allow

$$\mu_f'(x_2) \le (y(w, b) - y(a, v))^2, \qquad \forall x_2 \in [v, w] \subseteq [a, b].$$

4.1.1. Two-level case

The quantization error need not be balanced for a local or global minimum or maximum $x_{2,0}$ meaning that possibly $\int_a^{x_{2,0}} (f(u) - y_1(x_{2,0}))^2\,du \ne \int_{x_{2,0}}^b (f(u) - y_2(x_{2,0}))^2\,du$. The partitions derived by **FP** are given for $n = 3$ and invertible function f by a sequence of successive approximations $x_2^{(k+1)} = g_f(x_2^{(k)})$ of the associated fixed-point function

$$g_f(x) = f^{-1}\left(\frac{y_1(x) + y_2(x)}{2}\right) = f^{-1}\left(\frac{(F(x)/x - a) + (1 - F(x)/b - x)}{2}\right),$$

with level functions $y_1(x) := y(a, x)$ and $y_2(x) := y(x, b)$. As function f is increasing, the level functions y_1 and y_2 are increasing and hence, f^{-1} and g_f are. This implies that the argument sequence $(x_2^{(k)})_{k=1}^{\infty}$ is either increasing or decreasing and both of the sequences $(y_1(x_2^{(k)}))_{k=1}^{\infty}$ and $(y_1(x_2^{(k)}))_{k=1}^{\infty}$ are monotone in the same direction as the argument sequence.

LEMMA 14. *For twice differentiable (not necessarily monotone) function f the level functions y_1 and y_2 are both convex, if f is and they are both concave for a concave function f.*

PROOF. The level functions are twice differentiable with

$$y_1''(x) = \frac{f'(x)}{x-a} - \frac{2f(x)}{(x-a)^2} + \frac{2F(x)}{(x-a)^3} \qquad \text{and} \qquad y_2''(x) = 2\frac{1-F(x)}{(b-x)^3} - 2\frac{f(x)}{(b-x)^2} - \frac{f'(x)}{b-x}.$$

For convex f the level function y_1 is convex, since $\delta(x) := (x-a)^3 y_1''(x) \geq 0$ for all $x \in [a,b]$, which follows from $\delta(a) = 0$ and $\delta'(x) = f''(x)(x-a)^2 \geq 0$ for all $x \in [a,b]$. Function y_2 being convex and the case of concave f are treated in the analogous way. ∎

For quadratic f even the difference functions $f(x) - y_1(x)$ and $f(x) - y_2(x)$ are convex if f is convex and they are concave if f if so. It is worth noticing that the objective of the quantization problem itself, i.e., $\mu_f(x_2)$ need not be convex in x_2, even if f is increasing and strict convex.

EXAMPLE 1. For $D = [0,1]$ and $f(x) = 3x^2$ the unique optimum quantization is $x_2^0 = (1 + \sqrt{17})/8$. The objective μ_f is convex on $[-1/4 + \sqrt{1/16 + 1/6}, 1]$ but concave on $[0, -1/4 + \sqrt{1/16 + 1/6}]$. The inflexion point between the convex and concave segments is not a saddle point. However, μ is quasiconvex on $[0,1]$—i.e., $\mu(\alpha x + (1-\alpha)y) \leq \max\{\mu(x), \mu(y)\}$ for $\alpha \in [0,1]$—which guarantees the strict local minimum to be globally minimal. The fixed-point function $g_f(x) = \sqrt{(1 + x + 2x^2)/6}$ is convex over D. ∎

The fixed-point function g_f need not be convex even if f is and g_f can have several fixed points if f is increasing but neither convex nor concave.

EXAMPLE 2. Let $D = [0, 12]$ and

$$f(x) := \begin{cases} 10x, & \text{if } 0 \leq x \leq 1, \\ \dfrac{1}{10} \cdot x + \dfrac{99}{10}, & \text{if } 1 \leq x \leq 11, \\ 10x - 99, & \text{if } 11 \leq x \leq 12. \end{cases}$$

The value $\bar{x} = 6$ is a fixed point of g_f with $y_1(\bar{x}) = 75/8$ and $y_2(\bar{x}) = 93/8$. For $x^{(1)} = 1$ the sequence $(x^{(1)}, g_f(x^{(1)}), \ldots)$ is decreasing and bounded. For $x^{(1)} = 11$ the sequence is bounded and increasing. The limits are thus distinct and are fixed points of g_f since g_f is continuous. ∎

An observation from Example 2 is that each fixed point lies on a different linear segment. This is part of a general pattern.

THEOREM 3. *For an increasing spline f with polynomial of degree m over some segment of $[a,b]$ the function g_f has at most m fixed points over that segment except when a or b belongs to that segment where the bound increases to $m+1$. A fixed point in a or b is half-sided repelling.*

PROOF. As f is polynomial of degree m over some segment, F is polynomial of degree $m+1$ on that segment so that the fixed-point characterization

$$2f(x) = y_1(x) + y_2(x) \iff 2f(x)(x-a)(b-x) = F(x)(b-x) + (1-F(x))(x-a), \qquad x \in (a,b)$$

amounts to a polynomial equation of degree $m+2$. As $x = a$ and $x = b$ are solutions—which both do not belong to interior segments—there are at most m other solutions. ∎

Nonalgebraic functions can lead to a continuum of fixed points as follows. The fixed-point relation $g_f(x) = x$ holding for all $x \in [c,d] \subseteq (a,b)$ is equivalent to the integral function $F(x) = \int_a^x f(u)\,du$ satisfying the inhomogenous linear differential equation

$$F'(x) + F(x)\left[\frac{1}{2(b-x)} - \frac{1}{2(x-a)}\right] = F(b)\frac{1}{2(b-x)}, \qquad \forall x \in [c,d].$$

The general solution is $F(x) = C \cdot \sqrt{b-x}\sqrt{x-a} + (x-a)/(b-a) \cdot F(b)$, $C \in \mathbb{R}$ with $f(x) = F'(x) = C/2 \cdot (a+b-2x)/(\sqrt{b-x}\sqrt{x-a}) + F(b)/(b-a)$ where $f'(x) \geq 0$, $\forall x \in (a,b)$ if $C < 0$

and $f'(x) \leq 0$, $\forall x \in (a,b)$ if $C > 0$. As $F'(x)$ cannot be defined continuously for a and b there cannot be an interval of fixed points containing the boundary of the domain of f. A proper closed subinterval of (a,b), however, can and the optimal value function μ_f is constant over that subinterval.

EXAMPLE 3. Let $D = [a,b] = [0,10]$, $[c,d] = [8,9]$, $C = -2$, and $F(b) = 20$. The nonnegative function

$$f(x) = F'(x) := \begin{cases} g_1(x), & x \in [0,8], \\ \dfrac{(2x-10)}{(\sqrt{10-x}\,\sqrt{x})} + 2, & x \in [8,9], \\ g_2(x), & x \in [9,10], \end{cases}$$

is set to be continuous and increasing over $[0,10]$ by nonnegative functions $g_1(x)$ and $g_2(x)$ being continuous and increasing with $g_1(8) = 3.5$, $g_2(9) = 14/3$, $\int_0^8 g_1(u)\,du = \int_9^{10} g_2(u)\,du = 8$. This results in

$$F(x) = \begin{cases} \displaystyle\int_0^x g_1(u)\,du, & x \in [0,8], \\ -2\sqrt{10-x}\,\sqrt{x} + 2x, & x \in [8,9], \\ \displaystyle 12 + \int_9^x g_2(u)\,du, & x \in [9,10], \end{cases}$$

with $F(8) = 8$ and $F(9) = 12$. Each point from the interval $[8,9]$ is a fixed point of g_f. ∎

The functions g_1 and g_2 from the previous example can both be chosen to be also convex with $g_1'(8) = f'(8) = 50/64$ and $g_2'(9) = f'(9) = 50/27$. Thus, there exist functions f which are convex and strictly increasing over all D and yielding a continuum of fixed points with associated fixed-point function g_f being not convex over all D. Moreover, an increasing function with an arbitrary number of disjoint regions of fixed points can be constructed. The function must therefore be a solution of the inhomogenous differential equation on theses regions and different from that solution but increasing outside these regions.

DEFINITION 2. *A continuous function f whose associated fixed-point function g_f has finitely many fixed points only is called a normal function (with respect to fixed-point behaviour).*

From now on only normal functions will be considered. Normal functions are not of the form $k_1(a+b-2x)/(\sqrt{b-x}\,\sqrt{x-a}) + k_2$ on any interval $[c,d] \subseteq (a,b)$. Graphical analysis quickly reveals that a normal function has an odd number of fixed points where counting is according to multiplicity. The fixed points can be classified according to their attracting behaviour.

THEOREM 4. CHARACTERIZATION OF FIXED POINTS. *For monotone f each local minimum of μ_f is a contracting fixed point of g_f, each local maximum is a repelling fixed point, and each saddle point is half-sided attracting and half-sided repelling.*

PROOF. Only the case of a local minimum is considered. The contraction property of a fixed point implies local minimality, see Lemma 12.

Let $x_{2,0}$ be a local minimum of μ_f. Then $\mu_f'(x_{2,0}) = 0$, and hence, $g_f(x_{2,0}) = x_{2,0}$. Assuming $x_{2,0}$ were repelling implies that for each neighborhood $U(x_{2,0})$—small enough to contain no other fixed point of g_f—there is $x^{(1)} \in U(x_{2,0})$ such that $(x^{(k)})_{k=1}^{\infty}$ does not converge to $x_{2,0}$. Let $x^{(1)} > x_{2,0}$ without loss of generality.

The sequence $(x^{(k)})_{k=1}^{\infty}$ is increasing because if it were decreasing (no other situation possible as g_f is increasing) it would converge towards the unique fixed point $x_{2,0} \in U(x_{2,0})$. It converges to the smallest fixed point $x_{2,1} > x_{2,0}$. The point $x_{2,1}$ also is a critical point of μ_f, see Lemma 11. This next critical point can thus be reached by a sequence of decreasing values μ_f. Point $x_{2,0}$ being a local minimum then implies that there is a critical point (a local maximum) between $x_{2,0}$ and $x_{2,1}$, a contradiction. ∎

Fixed-point analysis and Lipschitz bounding combined result in a branch and bound strategy for the globally optimal quantization. The bounds are derived from a so-called saw-tooth argument which is based on two linear minorants. If μ_f is known to have Lipschitz constant λ over some interval $[c,d] \subseteq [a,b]$, then [15, Formula (16), p. 598]

$$\mu_f(x) \geq \frac{\mu_f(c)+\mu_f(d)}{2} - \lambda \frac{d-c}{2}, \qquad \forall x \in [c,d].$$

DEFINITION 3. *A subinterval of $[a,b]$ will be labeled explored whenever*

- *one endpoint is reached by a sequence of successive approximations initialized with the other endpoint, or*
- *the sequence of successive approximations initialized with one endpoint leaves the subinterval, or*
- *the lower bound of μ_f over that subinterval exceeds the minimal value of μ_f found so far.*

This notion comprises the trivial case of a singleton to be labeled *explored* if this singleton should happen to be a fixed point used as initial value to the successive approximations. The branching strategy may amount to merely choosing the mean of the boundary values of an unexplored interval as initial value $x^{(1)}$ to the sequence of successive approximations. If this sequence converges to a value not reached as a limit before, a new local minimum has been found. A potentially better branching rule for an unexplored interval $[c,d]$ is to choose the point where the maximum of the two linear minorants attains its minimum

$$x^{(1)} = \frac{c+d}{2} + \frac{\mu_f(c)-\mu_f(d)}{2\lambda}.$$

GL1

1. Initialization. The values $x^{(1)} = a$ and $x^{(1)} = b$ are chosen as initial values to the successive approximations $x^{(k+1)} = g_f(x^{(k)})$ resulting in the limits z_1 and z_2, respectively. Intervals $[a,z_1]$ and $[z_2,b]$ are the only labeled *explored*. $F \leftarrow \min\{\mu_f(z_1), \mu_f(z_2)\}$, $Li \leftarrow \{\arg\min\{\mu_f(z_1), \mu_f(z_2)\}\}$.
2. Iteration. As long as there is a subinterval of $[a,b]$ without a label *explored* do
   (a) Initialization of **FP**. A value $x^{(1)}$ of such a subinterval is chosen as initial value for the successive approximations of **FP**.
   Termination analysis of **FP**.
   i. If the successive approximations converge to a value z not reached as limit before, then interval $[\min\{x^{(1)},z\}, \max\{x^{(1)},z\}]$ receives label *explored*. If $f(z) < F$, then $F \leftarrow f(z)$ and $Li \leftarrow \{z\}$. If $f(z) = F$, then $Li \leftarrow Li \cup \{z\}$.
   ii. If the successive approximations enter an *explored* interval by crossing its boundary δ, then the interval $[\min\{x^{(1)},\delta\}, \max\{x^{(1)},\delta\}]$ receives label *explored*.
   (b) Intervals without label *explored* are tested for receiving this label due to a decrease in F or a reduction in the length of some interval.

Case 2(a)i. comprises that of a constant sequence of successive approximations. The procedure **FP** is called finitely many times only by **GL1** as each of the finite many local minima is eventually reached by a decreasing or increasing sequence of successive approximations. This does not violate the fact that global minima of Lipschitz bounded functions can generally not be found by algorithms using only a finite number of function evaluations [15, p. 589] since a call of **FP** requires to compute level functions of f and it may—conceptually—compute an infinite sequence of function evaluations.

4.1.2. Two and more levels

The sequences of levels $(y_i^{(k)})_{k=1}^{\infty}$ generated by **FP** need neither be increasing nor decreasing in case of $n-1 > 2$ levels. The general structure of attracting domains of local minima of

the quantization objective cannot be fully described. However, inner approximations can be given by unions of intervals and in low degree cases relative locations of local minima can be described by the componentwise order. This is defined on the Euclidean space $\mathbb{R}^N$ as usual by $x \leq y :\Longleftrightarrow (x_1 \leq y_1) \ldots (x_N \leq y_N)$.

LEMMA 15. *Any two local extrema of μ_f are comparable in the componentwise order for $n = 4$ and $n = 5$.*

PROOF. Only for $n = 4$ as the argument also applies to $n = 5$. Let $(x'_{2,0}, x'_{3,0})$ and $(x''_{2,0}, x''_{3,0})$ be local extrema of μ_f. The restricted problems $\min_{x_3} \mu_f(x_{2,0}, x_3)$ and $\max_{x_3} \mu_f(x_{2,0}, x_3)$, respectively, then have solutions at $x_3 = x'_{3,0}$ if $x_{2,0} = x'_{2,0}$ and at $x_3 = x''_{3,0}$ if $x_{2,0} = x''_{2,0}$. Hence,

$$\frac{\partial \mu_f(x_{2,0}, x_3)}{\partial x_3} = 0 \Longleftrightarrow x_3 = f^{-1}\left(\frac{y(x_{2,0}, x_3) + y(x_3, b)}{2}\right).$$

A solution of the last equation is increasing in values for $x_{2,0}$. ∎

The componentwise order is preserved by successive approximations.

LEMMA 16. *Let two sequences $(x^{(k)})_{k=1}^{\infty}$ and $(v^{(k)})_{k=1}^{\infty}$ be given over Δ_n with $x^{(k+1)} = g_f(x^{(k)})$ and $v^{(k+1)} = g_f(v^{(k)})$, $\forall k \in \mathbb{N}$.*

1. *The inequality $x^{(1)} \leq v^{(1)}$ implies $x^{(k)} \leq v^{(k)}$, $\forall\, k \in \mathbb{N}$.*
2. *The inequality $x^{(1)} \geq v^{(1)}$ implies $x^{(k)} \geq v^{(k)}$, $\forall\, k \in \mathbb{N}$.*

PROOF. Part 1 is shown only by induction on k. Monotonicity of f and monotonicity of y in both arguments result for all coordinates $i \in \{2, \ldots, n-1\}$ in

$$\begin{aligned} v_i^{(k+1)} &= f^{-1}\left(\frac{y\left(v_{i-1}^{(k)}, v_i^{(k)}\right) + y\left(v_i^{(k)}, v_{i+1}^{(k)}\right)}{2}\right) \\ &\geq f^{-1}\left(\frac{y\left(x_{i-1}^{(k)}, x_i^{(k)}\right) + y\left(x_i^{(k)}, x_{i+1}^{(k)}\right)}{2}\right) = x_i^{(k+1)}. \end{aligned}$$

∎

As a consequence of the previous result, certain intervals can be ruled out to have more than one fixed point of g_f.

LEMMA 17. *Let $z = \lim_{k\to\infty} x^{(k)}$.*

1. *For $x^{(1)} \leq z$ the interval $[x^{(1)}, z]$ contains no fixed point of g_f except z.*
2. *For $x^{(1)} \geq z$ the interval $[z, x^{(1)}]$ contains no fixed point of g_f except z.*

PROOF. Part 1 need to be shown only. Assuming the existence of a fixed point $v \in [x^{(1)}, z] - \{z\}$ implies $x^{(k)} \leq v \leq z$ by choosing $v^{(1)} = v = v^{(k)}$ according to Lemma 16. This contradicts the convergence $x^{(k)} \to z$ $(k \to \infty)$ because $v_i < z_i$ for at least one coordinate $i \in \{2, \ldots, n-1\}$. ∎

The previous results for $x^{(1)} \leq z = \lim_{k\to\infty} x^{(k)}$ neither imply that $x^{(k)} \leq x^{(k+1)}$, $\forall\, k \in \mathbb{N}$ nor that the sequence $(x^{(k)})_{k=1}^{\infty}$ is contained in $[x^{(1)}, z]$. Both statements are wrong in general where the latter may be true for a nondegenerate interval $[x^{(1)}, z]$, i.e., for $\operatorname{int}([x^{(1)}, z]) \neq \emptyset$. The situation of the sequence $(x^{(k)})_{k=1}^{\infty}$ leaving the interval $[x^{(1)}, z]$ and still having limit value z is different from the two-level case with one-dimensional fixed-point functions.

In case of more than two quantization levels, the difference between function values and the average of adjacent levels does generally not decrease with the number of iterations. For each function f there is a vector $x^{(k)}$ such that for at least one coordinate $i \in \{2, \ldots, n-1\}$

$$\left| f\left(x_i^{(k+1)}\right) - \frac{y\left(x_{i-1}^{(k+1)}, x_i^{(k+1)}\right) + y\left(x_i^{(k+1)}, x_{i+1}^{(k+1)}\right)}{2} \right|$$

$$\not\leq \left| f\left(x_i^{(k)}\right) - \frac{y\left(x_{i-1}^{(k)}, x_i^{(k)}\right) + y\left(x_i^{(k)}, x_{i+1}^{(k)}\right)}{2} \right|.$$

The componentwise ordering of local minima in cases $n = 4$ and $n = 5$ allows a globally convergent branch and bound procedure in analogy to **GL1**. The basic idea is to use initial values for successive approximations which componentwise lie between previously detected local minima of the quantization objective.

The notion of exploration is extended to multidimensional intervals $[c, d]$ which have the two and only two componentwise extreme points c and d. A subinterval $[c, d]$ of $[a, b] \subseteq \mathbb{R}^{n-2}$ is called *explored* if

- one componentwise extremal endpoint is reached by successive approximations initialized with the other extremal endpoint or
- the sequence of successive approximations initialized with one extremal endpoint enters an *explored* subinterval or
- the lower bound of μ_f over that subinterval exceeds the minimal value of μ_f found so far.

The quantization algorithm for multiple levels differs from the two-level case by an interval being declared *explored* according to the second case of the definition even if this interval has not been traversed from one extremal endpoint to the other. Suppose the sequence $(x^{(k)})_{k=1}^{\infty}$ with $x^{(1)} \in [z_1, z_2]$ runs into the interval $[w^{(1)}, z_2]$ where $\lim_{k\to\infty} w^{(k)} = z_2$ as in Figure 3. The whole interval $[x^{(1)}, z_2]$, respectively, the whole interval $[x^{(1)}, \omega]$ cannot contain a fixed point of g_f according to Lemma 17 and is hence, labeled *explored*. It even follows that $[x^{(1)}, \omega] \subseteq \mathrm{dom}(z_2)$.

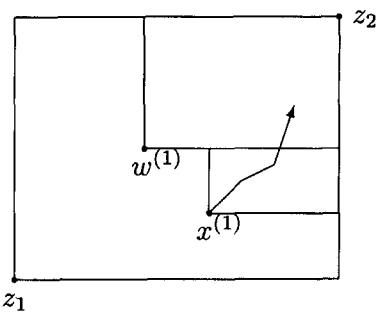

Figure 3. $x^{(k)}$ running into $\mathrm{dom}(z_2)$.

The quantization objective has the two minorants $\varphi_1(x) = \mu_f(c) - \lambda\|x - c\|$ and $\varphi_2(x) = \mu_f(d) - \lambda\|x - d\|$ over a subinterval $[c, d]$. The minimum value $\min_{x\in[c,d]} \varphi(x)$ of the lower bound $\varphi(x) = \max\{\varphi_1(x), \varphi_2(x)\}$ is hard to compute for multidimensional intervals $[c, d]$.

The minimum of φ along the main diagonal of $[c, d]$ is attained at [15, Formula (40), p. 607]

$$x^{(1)} = \frac{c+d}{2} + (d-c)\,\frac{\mu_f(c) - \mu_f(d)}{2\lambda\,\|d-c\|}.$$

This value may serve for the branching part of the subsequent global procedure for $n = 4, 5$.

GL2 Small

1. Initialization. The vectors $x^{(1)} = (a, \ldots, a)$ and $x^{(1)} = (b, \ldots, b)$ are chosen as initial values for $x^{(k+1)} = g_f(x^{(k)})$ resulting in the limits z_1 and z_2, respectively. Consideration is restricted to $[z_1, z_2]$ and $[z_1, z_2]$ is labeled *explored* if and only if $z_1 = z_2$. $F \leftarrow \min\{\mu_f(z_1), \mu_f(z_2)\}$, $Li \leftarrow \{\arg\min\{\mu_f(z_1), \mu_f(z_2)\}\}$.
2. Iteration. As long as there is a subinterval $[c, d]$ of $[z_1, z_2]$ without a label *explored* do
   (a) Initialization of **FP**. A vector $x^{(1)} \in [c, d] \cap \Delta([a, b])$ is chosen as initial value for the successive approximations of **FP**.
   Termination analysis of **FP**.
   i. If the successive approximations converge to a value z not reached as limit before with z_a reached as largest lower and z_b reached as smallest upper limit of z before,

then consideration in $[z_a, z_b]$ is restricted to $[z_a, z] \cup [z, z_b]$. Thus, all intervals $I \subseteq [z_a, z]^c \cap [z, z_b]^c$ are labeled *explored*; complementation refers to $[z_a, z_b]$. If $x^{(1)} \leq z$, then also the interval $[x^{(1)}, z]$ is labeled *explored* and if $x^{(1)} \geq z$, then also the interval $[z, x^{(1)}]$ is labeled *explored*. If $f(z) < F$, then $F \leftarrow f(z)$ and $Li \leftarrow \{z\}$. If $f(z) = F$, then $Li \leftarrow Li \cup \{z\}$.

ii. If the successive approximations converge to a value z reached as limit before or enter an *explored* interval $\subseteq \mathrm{dom}(z)$ then $[x^{(1)}, z]$ is labeled *explored* if $x^{(1)} \leq z$ and $[z, x^{(1)}]$ is labeled *explored* if $z \leq x^{(1)}$.

(b) Intervals in $[z_1, z_2]$ without label *explored* are tested for receiving this label.

In Step 2(a)i. neither $x^{(1)} \leq z$ nor $x^{(1)} \geq z$ may occur.

LEMMA 18. *Any two incomparable fixed points v, w of g_f are sandwiched by local minima z_1, z_2 of μ_f, i.e., $z_1 \leq v, w \leq z_2$.*

PROOF. SKETCH. Choosing $x^{(1)} = (\max\{v_2, w_2\}, \ldots, \max\{v_{n-1}, w_{n-1}\})$ results in an increasing sequence $x^{(1)} \leq x^{(2)} = g_f(x^{(1)}) \leq \ldots$ with $v, w \leq \lim_{k\to\infty} x^{(k)} =: z_2$. The analog applies to $x^{(1)} = (\min\{v_2, w_2\}, \ldots, \min\{v_{n-1}, w_{n-1}\})$ resulting in $z_1 := \lim_{k\to\infty} x^{(k)}$. ∎

The restriction in Step 2(a)i of **GL2 small** is infeasible in the case $n \geq 6$. In that case, the global minimization of μ_f goes along a modification of the algorithm by Piavskii and Pinter [15]. All local minima lie between the smallest local minimum z_1 and the largest local minimum z_2 as computed in Step 1 of **GL2 small**. Subdivision of intervals is along hyperplanes which are orthogonal to one axis. The search for a local minimum within such an interval $[c, d]$ can be reduced, if $c, g_f(c), \ldots$ converges to $c' \in [c, d]$ or if $d, g_f(d), \ldots$ converges to $d' \in [c, d]$; $c \leq d$ implies $c' \leq d'$ and $\min_{x\in[c,d]} \mu_f(x) = \min_{x\in[c',d']} \mu_f(x)$. Furthermore, a lower bound of μ_f over $[c, d]$ with $z \in [c, d]$ is called improved bound and given by

$$\min_{x\in[c,d]} \mu_f(x) \geq \max\{\mu_f(c) - \lambda\|c - d\|, \mu_f(d) - \lambda\|c - d\|, \mu_f(z) - \lambda \max\{\|c - z\|, \|d - z\|\}\}.$$

A subdivision of $[c, d]$ by a hyperplane orthogonal to one axis and through $z \in [c, d]$ will be chosen orthogonal to an axis with largest difference of coordinates, i.e., according to $i_0 = \arg\max_{i=2,\ldots,n-1}\{d_i - c_i\}$. This tends to decrease the lengths of the main diagonals of subsequent subintervals and hence, tends to improve future lower bounds for μ_f.

GL2 Large

1. Initialization. The vectors $x^{(1)} = (a, \ldots, a)$ and $x^{(1)} = (b, \ldots, b)$ are chosen as initial values for $x^{(k+1)} = g_f(x^{(k)})$ resulting in the limits z_1 and z_2, respectively. Consideration is restricted to the interval $[z_1, z_2]$ which is labeled *explored* if and only if $z_1 = z_2$. $F \leftarrow \min\{\mu_f(z_1), \mu_f(z_2)\}$, $Li \leftarrow \{\arg\min\{\mu_f(z_1), \mu_f(z_2)\}\}$.
2. Iteration. As long as there is a subinterval $[c, d]$ of $[z_1, z_2]$ without a label *explored* do
   (a) Initialization of **FP**. A vector $x^{(1)} \in \mathrm{conv}\{c, d\}$ is chosen as initial value for the successive approximations of **FP** with $\lim_{k\to\infty} x^{(k)} = z$.

   Termination analysis of **FP**.

   i. If $z \in [c, d]$ and if z was reached as limit (with other initial value $x^{(1)}$) before, then $[c, d]$ is devided by a hyperplane orthogonal to one axis and through $x^{(1)}$.

   ii. If $z \in [c, d]$ and if z was not reached as limit before, then $[c, d]$ is devided by a hyperplane orthogonal to one axis and through z. If $f(z) < F$, then $F \leftarrow f(z)$ and $Li \leftarrow \{z\}$. If $f(z) = F$, then $Li \leftarrow Li \cup \{z\}$.

   iii. If $z \notin [c, d]$ and if each earlier obtained limit $z_0 \in [c, d]$ lies on an earlier established hyperplane, then $[c, d]$ is devided by a hyperplane orthogonal to one axis and through $x^{(1)}$.

iv. If $z \notin [c,d]$ and if an earlier obtained limit $z_0 \in [c,d]$ does not lie on any earlier established hyperplane, then $[c,d]$ is devided by a hyperplane orthogonal to one axis and through z_0.

(b) Intervals in $[z_1,z_2]$ without label *explored* are tested for receiving this label.

4.2. Nonmonotone Functions

For algorithmic purposes the domain of a nonmonotone function is segmented into $\subseteq$-maximal monotonicity intervals $[a_m,b_m]$ of which a finite number M is admitted. There are again assumed to be no constant line segments and ranges of the monotonicity segments are bounded by $l_m := \inf_{x\in[a_m,b_m]} f(x)$ and $u_m := \sup_{x\in[a_m,b_m]} f(x)$. The monotone parts of f are $f_m : [a_m,b_m] \to [l_m,u_m]$ with $f_m(x) = f(x)$ for all $x \in [a_m,b_m]$, and f_m^{-1} exists over $[l_m,u_m]$, $m = 1,\ldots,M$.

For each $(x_2,\ldots,x_{n-1}) \in \Delta([a,b])$ there is at least one f_m attaining the value $(y(x_{i-1},x_i) + y(x_i,x_{i+1}))/2$. The general procedure **GenP** specializes to the subsequent fixed-point procedure **FPmonseg** searching in a particular monotonicity segment of f as long as possible.

FPmonseg

1. Initialization. An arbitrary partition $a = x_1^{(1)} \le x_2^{(1)} \le \cdots \le x_{n-1}^{(1)} \le x_n^{(1)} = b$ is selected. Set $k \leftarrow 1$.
2. Iteration. Repetition until some stopping criterion is met:
   (a) Computation of levels $y_i^{(k)} = y(x_i^{(k)}, x_{i+1}^{(k)})$, $i = 1,\ldots,n-1$.
   (b) Computation of a new partition with $x_i^{(k+1)} = g_{f,m}(x_i^{(k)}) = f_m^{-1}((y_{i-1}^{(k)} + y_i^{(k)})/2)$ with valid sign condition $\operatorname{sgn} f'(x_i^{(k+1)}) = \operatorname{sgn}(y_i^{(k)} - y_{i-1}^{(k)})$ and $x_i^{(k+1)} \in [a_m,b_m]$ where $x_i^{(k)} \in [a_m,b_m]$ with $m = m(i)$, $i = 2,\ldots,n-1$. If no such $x_i^{(k+1)}$ can be found in $[a_m,b_m]$, another monotonicity segment is chosen. (Monotonicty of f_m over $[a_m,b_m]$ implies that there is at most one candidate $x_i^{(k+1)} \in [a_m,b_m]$.)
   (c) $k \leftarrow k+1$.

REMARK 3. For nonmonotone f

1. the levels $y_i^{(k)}$ need neither be arranged increasingly nor decreasingly for fixed k, and
2. it is even possible that $f(x_i^{(k)}) \notin [\min\{y_{i-1}^{(k)}, y_i^{(k)}\}, \max\{y_{i-1}^{(k)}, y_i^{(k)}\}]$.

4.2.1. Two-level case

For nonmonotone f the level functions y_1 and y_2 need not be of consistent monotonicity behaviour on a segment where f is. In fact, both functions y_1 and y_2 may be decreasing on a segment where f is increasing. This makes it difficult to rule out or confirm "*a priori*" that a local minimum of μ_f lies in that segment. However, if monotonicity of the level functions is present on boundaries of the segment, it can be extended into the interior.

LEMMA 19. *Let f be differentiable on $[a,b]$.*

1. *If the monotone part f_m is increasing, then $y_1'(a_m) \ge 0 \Longrightarrow y_1'(x) \ge 0$, $\forall x \in [a_m,b_m]$ and $y_2'(b_m) \ge 0 \Longrightarrow y_2'(x) \ge 0$, $\forall x \in [a_m,b_m]$.*
2. *If the monotone part f_m is decreasing, then $y_1'(a_m) \le 0 \Longrightarrow y_1'(x) \le 0$, $\forall x \in [a_m,b_m]$ and $y_2'(b_m) \le 0 \Longrightarrow y_2'(x) \le 0$, $\forall x \in [a_m,b_m]$.*

PROOF. Only for y_1 and increasing f_m. Considering function $\eta(x) := (x-a)^2 y_1'(x) = (x-a)f(x) - F(x)$ implies the desired nonnegativity as $\eta'(x) = (x-a)f'(x) \ge 0$ and $\eta(a_m) = (a_m-a)^2 y_1'(a_m) \ge 0$. ∎

DEFINITION 4. *The closed set*

$$k_f([a_m, b_m]) := k([a_m, b_m]) := \left\{ x \mid x \in [a_m, b_m] \text{ with } l_m \leq \frac{(y_1(x) + y_2(x))}{2} \leq u_m \right\}$$

is denoted as the kernel of $[a_m, b_m]$ *with respect to function* f.

For all $x \in k_f([a_m, b_m])$ the local fixed-point function

$$g_{f,m}(x) := f_m^{-1}\left(\frac{y_1(x) + y_2(x)}{2}\right)$$

is defined with values in $[a_m, b_m]$. All local fixed-point functions are summarized as the fixed-point function $g_f : \bigcup_{m=1}^{M} k_f([a_m, b_m]) \to [a, b]$ with $g_f(x) = g_{f,m}(x)$, $\forall x \in k_f([a_m, b_m])$. Thus, a slight incorrectness is accepted by allowing g_f to attain two values at points which belong to two adjacent kernels. Ranges of g_f over different kernels are disjoint. The local fixed-point functions are also called branches. Branches need not be monotone.

EXAMPLE 4. Let f be given over $D = [0, 11]$ by

$$f(x) := \begin{cases} -10x + 10, & x \in [0, 1] = [a_1, b_1], \\ x - 1, & x \in [1, 11] = [a_2, b_2]. \end{cases}$$

As $[l_1, u_1] = [l_2, u_2] = [0, 10]$, both branches are defined over their complete monotonicity segments with level functions

$$y_1(x) = \begin{cases} 10 - 5x, & 0 \leq x \leq 1, \\ \dfrac{x}{2} - \dfrac{1 + 5.5}{x}, & 1 \leq x \leq 11, \end{cases} \quad \text{and} \quad y_2(x) = \begin{cases} 5 \cdot \dfrac{(x^2 - 2x + 11)}{(11 - x)}, & 0 \leq x \leq 1, \\ \dfrac{x}{2} + \dfrac{9}{2}, & 1 \leq x \leq 11. \end{cases}$$

This leads to $g_{f,1}(1) = .5$ and $g_{f,2}(1) = 6$. Branch $g_{f,1}$ is increasing over $[0, 1]$. The sum of the level functions $y_1(x) + y_2(x)$ has a critical point at $\sqrt{5.5}$ which results in $g_{f,2}$ being decreasing over $[1, \sqrt{5.5}]$ and increasing over $[\sqrt{5.5}, 11]$. ∎

The kernels of all but one monotonicity segment may be empty but the union of all kernels is not. Considerable difficulties are caused by nonvoid kernels not containing their boundaries, i.e., $\emptyset \neq k_f([a_m, b_m]) \subseteq (a_m, b_m)$. Such kernels exist.

EXAMPLE 5. Let f be given over $D = [0, 4]$ with parameter $\gamma > 8$ and

$$f(x) := \begin{cases} -\gamma x + 2\gamma, & x \in [0, 1] = [a_1, b_1], \\ x + \gamma - 1, & x \in [1, 3] = [a_2, b_2], \\ -\gamma x + 4\gamma + 2, & x \in [3, 4] = [a_3, b_3]. \end{cases}$$

The outer segments coincide with their kernels. The middle segment satisfies $\emptyset \neq k_f([a_2, b_2]) \subseteq (a_2, b_2)$ since $(y_1(2) + y_2(2))/2 = \gamma + 1 \in [\gamma, \gamma + 2] = [l_2, u_2]$, but $(y_1(1) + y_2(1))/2 = 7/6 \cdot \gamma + 2/3 > \gamma + 2$ and $(y_1(3) + y_2(3))/2 = 5/6 \cdot \gamma + 4/3 < \gamma$.

The branches $g_{f,1}$ and $g_{f,3}$ are increasing over their domains while $g_{f,2}$ is decreasing. ∎

The fixed-point function g_f for a nonmonotone f may thus be only partially defined over $D = [a, b]$. If it is defined in the neighbourhood of a change of monotonicity point a_m, then the fixed-point function has a jump discontinuity in a_m—apart from the fact that it is doubly defined there. The kernel of a monotonicity segment cannot be ensured to be a connected set. Deciding whether a monotonicity segment has a nonvoid kernel and if so, computing its boundary is not trivial. None of the change of monotonicity points is a local minimum of μ_f which is revealed

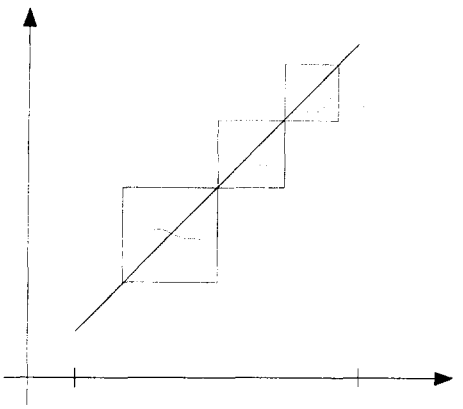

Figure 4. Possible appearance of fixed-point function.

by a geometric distortion argument. Figure 4 shows a possible appearance of the fixed-point function. Function g_f is guaranteed to have a fixed point in $[a_m, b_m]$ if $\{a_m, b_m\} \subseteq k_f([a_m, b_m])$.

Fixed points of g_f correspond to critical points of μ_f but μ_f can have "more" critical points. Such are local maxima of μ_f with equal levels meaning that they satisfy the condition $y_1(x_{2,0}) = y_2(x_{2,0})$. A local maximum of μ_f with equal levels is called obvious local maximum and it may or may not correspond to a fixed point.

LEMMA 20. *Each fixed point of $g_{f,m}$ is a critical point of μ_f and each critical point $x_{2,0} \in [a_m, b_m]$ of μ_f which is not an obvious local maximum is a fixed point of $g_{f,m}$.*

PROOF. The argument is the same as for Lemma 11 and it ensures that a critical point of μ_f which is not an obvious local maximum, belongs to the kernel of the corresponding monotonicity segment. ∎

THEOREM 5. *A local minimum of μ_f is a contracting fixed point of g_f and vice versa.*

PROOF. Only the contracting property remains to be shown according to Lemma 12 which also holds for nonmonotone f. Let $x_{2,0}$ be a local minimum of μ_f. Then $g_f(x_{2,0}) = x_{2,0}$. Assume $x_{2,0}$ were not contracting. Then for each sufficiently small neighbourhood $U(x_{2,0})$ there is $x^{(1)} \in U(x_{2,0}) - \{x_{2,0}\}$ such that $(x^{(k)})_{k=1}^{\infty}$ does not converge to $x_{2,0}$ and $x_{2,0}$ is the only extreme point of μ_f in that neighbourhood.

As g_f is continuous, there is $x^{(1)} > x_{2,0}$ such that $x^{(3)} = g_f(g_f(x^{(1)})) \in U(x_{2,0})$. Then $x^{(3)} > x_{2,0}$ irrespective of the local direction of monotonicity of g_f. The inequality $x^{(3)} > x^{(1)}$ would result in a contradiction to the sequence $(\mu_f(x^{(k)}))_{k=1}^{\infty}$ being decreasing. Thus, $x_{2,0} < x^{(3)} < x^{(1)}$. This implies $x_{2,0} < x^{(k+1)} < x^{(k)}$ for each $x^{(k)}$ with $x^{(k)} > x_{2,0}$. The sequence $(x^{(k)})_{k=1}^{\infty}$ thus converges to $x_{2,0}$, a contradiction. ∎

The sign condition of Lemma 7 implies that two local minima of μ_f on adjacent monotonicity segments of f are separated by an obvious local maximum. The occurence of obvious local maxima is illustrated by Example 4. There, the objective μ_f has the critical point $x_{2,0} = 1$ since $y_1(1) = y_2(1) = 5$, but $x_{2,0} = 1$ is not a fixed point since $f(1) = 0 \neq (y_1(1) + y_2(1))/2$.

LEMMA 21. *The objective μ_f has at least one obvious local maximum for a nonmonotone function f with $f(a) = f(b) \neq 1/(b-a) \int_a^b f(u)\,du$.*

PROOF. The level functions y_1 and y_2 both intersect since $y_1(a) = f(a) = f(b) = y_2(b)$ and $y_1(b) = 1/(b-a) \int_a^b f(u)\,du = y_2(a)$. ∎

Though the sequence $x^{(1)}, g_{f,m}(x^{(1)}), \ldots$ has decreasing values $\mu_f(x^{(1)}) \geq \mu_f(g_{f,m}(x^{(1)})) \geq \ldots$, it may jump over local maxima of μ_f in $[a_m, b_m]$ for $x^{(1)} \in [a_m, b_m] - k_f([a_m, b_m])$. The sequence $x^{(1)}, g_{f,m}(x^{(1)}), \ldots$ may even jump over a local minimum of μ_f if initialized with $x^{(1)} \in [a, b] - [a_m, b_m]$. (The sequence of successive approximations may be well defined in such a case.) Thus, disconnected attracting domains of fixed points would exist, if local fixed-point functions

were considered on larger domains than their monotonicity segments. The search for fixed points within kernels can be reduced to those subsets where function f is sandwiched by the level functions.

DEFINITION 5. *For a nonmonotone function f with monotonicity segment $[a_m, b_m]$*

1. *the relevant kernel $rk_f([a_m, b_m])$ is defined to be*

$$rk_f([a_m, b_m]) := \Big\{ x \mid x \in [a_m, b_m] \text{ with } l_m \leq \frac{y_1(x) + y_2(x)}{2} \leq u_m \text{ and } \min\{y_1(x), y_2(x)\} \leq f(x) \leq \max\{y_1(x), y_2(x)\} \Big\}$$

2. *and the relevant segment is defined to be*

$$r_m([a_m, b_m]) := r_m := \{x \mid x \in [a_m, b_m] \text{ with } \min\{y_1(x), y_2(x)\} \leq f(x) \leq \max\{y_1(x), y_2(x)\}\}.$$

Obviously, $rk_f([a_m, b_m]) \subseteq r_m([a_m, b_m]) \subseteq [a_m, b_m]$ and $rk_f([a_m, b_m]) \subseteq k_f([a_m, b_m]) \subseteq [a_m, b_m]$, but there seems to be no general inclusion between $r_m([a_m, b_m])$ and $k_f([a_m, b_m])$. The relevant kernel of a function which is monotone over the whole interval $[a, b]$ coincides with that interval, i.e., $rk_f([a, b]) = [a, b]$. The analysis of relevant kernels benefits from a precise description of the monotonicity behaviour of the level functions. For later use level functions $y(x_i, \cdot)$ and $y(\cdot, x_i)$ are considered with arbitrary $x_i \in [a, b]$.

LEMMA 22. *All local optima of the level functions $y(x_i, x)$ with $x_i \leq x \leq b$ and $y(x, x_i)$ with $a \leq x \leq x_i$ coincide with intersection points with f where*

1. *level function $y(x_1, \cdot)$ has a local minimum, respectively, maximum if it intersects with f on an increasing, respectively, decreasing segment of f.*
2. *level function $y(\cdot, x_1)$ has a local maximum, respectively, minimum if it intersects with f on an increasing, respectively, decreasing segment of f.*

As a consequence, each level function has at most one intersection point with f on each monotonicity segment of f. Also, each level function has no more monotonicity segments than f. This is to be interpreted analogously to results on decreasing sign changes for polynomials and power series, see [16, Chapter 1, paragraph 3].

PROOF OF LEMMA 22.

PART 1. A critical point x_0 of the level function $y(x_i, x)$ satisfies $\frac{\partial^2 y(x_i,x)}{\partial^2 x}|_{x=x_0} = f'(x_0)/(x_0 - x_i) - 2\frac{\partial y(x_i,x)}{\partial x}|_{x=x_0}/(x_0 - x_i) = f'(x_0)/(x_0 - x_i)$ implying the result.

PART 2. Similarly as $\frac{\partial^2 y(x,x_i)}{\partial^2 x}|_{x=x_0} = 2\frac{\partial y(x,x_i)|_{x=x_0}}{(x_i - x_0)} - f'(x_0)/(x_i - x_0) = f'(x_0)/(x_i - x_0)$. ∎

THEOREM 6. *Let f consist of M monotonicity segments over $[a, b]$.*

1. *Both level functions y_1 and y_2 consistently are of the same monotonicity direction over each relevant segment $r_m([a_m, b_m])$.*
2. *All relevant kernels $rk_f([a_m, b_m])$ are either void or connected.*

PROOF.

PART 1. Without loss of generality f is assumed to be increasing over $[a_m, b_m]$.

CASE 1. Both level functions intersect with f over $[a_m, b_m]$. The relevant segment r_m is the interval between the intersection points and the level functions both are either increasing or decreasing there according to Lemma 22.

Further cases are given by level function y_2 having no intersection with f over $[a_m, b_m]$, i.e., level function y_2 is of constant monotonicity direction over segment $[a_m, b_m]$. Missing cases can be reduced to given ones by symmetry arguments.

CASE 2. Level function y_2 is increasing and below f. Level function y_2 properly intersects with f at some $x_0 > b_m$. If such an intersection point would not exist, then the rightmost monotonicity segment were increasing as sgn $f'(x) =$ sgn $y_2'(x)$, $\forall x \in [a_M, b_M]$ and y_2 is above f on $[a_M, b_M]$ for f and y_2 being increasing there.

The smallest proper intersection point $x_1 > b_m$ is a local maximum of y_2 since y_2 is increasing over $[a_m, b_m]$. The proper intersection then occurs over an increasing segment of f, see Lemma 22. This is impossible and hence, Case 2 cannot occur.

CASE 3. Level function y_2 is decreasing and below f. Let y_1 intersect with f at $x_0 \in [a_m, b_m]$. Then x_0 is a local minimum of y_1, and hence, y_1 and y_2 are both decreasing over $r_m = [a_m, x_0]$.

Assume y_1 does not intersect with f over $[a_m, b_m]$. If y_1 is decreasing, then $r_m = [a_m, b_m]$ or $r_m = \emptyset$. If y_1 were increasing, the same contradiction as in Case 2 would result.

CASE 4. Level function y_2 is increasing and above f. Let y_1 intersect with f at $x_0 \in [a_m, b_m]$. This corresponds to the intersection situation in Case 3 with now both y_1 and y_2 being increasing over $r_m = [x_0, b_m]$.

Assume y_1 does not intersect with f over $[a_m, b_m]$. If y_1 is increasing, then $r_m = [a_m, b_m]$ or $r_m = \emptyset$. If y_1 were decreasing, the corresponding contradiction from Case 3 would result.

CASE 5. Level function y_2 is decreasing and above f. Analog to Case 2.

PART 2. The function $(y_1(x) + y_2(x))/2$ is monotone over each relevant segment $r_m([a_m, b_m])$ since both level functions are. "Cutting-off" $r_m([a_m, b_m])$ according to lower and upper bounds of the monotone function $(y_1(x) + y_2(x))/2$ preserves the interval structure. ∎

The previous proof establishes that level functions y_1 and y_2 have identical monotonicity direction over each relevant segment. The first part of Theorem 6 implies a dichotomy of relevant segments and relevant kernels respectively into type **id** and type **diff**. A relevant segment or a relevant kernel is of type **id**, if function f and both level functions have identical monotonicity direction there. A relevant segment or a relevant kernel is of type **diff**, if function f and the level functions have different monotonicity directions there.

Computing the optimal quantization is thus, inherent to searching through type **id** and type **diff** segments. Type **diff** segments obviously have at most one local minimum of the quantization objective μ_f. A relevant type **diff** segment $[\alpha_m, \beta_m]$ can be checked easily for containing a fixed point of g_f by comparing the signs of $y_1(x) + y_2(x) - 2f(x)$ at $x = \alpha_m$ and $x = \beta_m$. The fixed point can be found by full iterated bisection (*regula falsi*) or by iterated bisection until a value of $rk_f([a_m, b_m])$ is reached and proceeded by successive approximation from there.

The computation of the relevant kernel $rk_f([a_m, b_m]) =: [\alpha_m', \beta_m']$ of a type **id** segment $[\alpha_m, \beta_m]$ is facilitated by a slight variant of *regula falsi*. The initial values therefore are α_m and β_m but search candidates are the left and right boundaries (not middle values) of the current interval. The calculation of relevant segments from segments requires at most to compute intersection points between f and the level functions. This can be done by *regula falsi* in principle. For computing the optimal quantization only relevant segments with satisfiable sign condition sgn $f'(x) =$ sgn $(y_2(x) - y_1(x))$ need to be considered, comp. Lemma 7. The satisfiability of the sign condition is easily established since the monotonicity type of f is known over each segment, and thus, over each relevant segment.

Local minima of the quantization objective μ_f can then be found by the global procedure **GL1** applied to relevant kernels of type **id**. Lipschitz bounds can be derived in a straightforward manner.

LEMMA 23. *Let $[\alpha_m, \beta_m]$ be a relevant segment or a relevant kernel of type* **id**. *The quantization objective $\mu_f(x)$ is then Lipschitz-bounded where*

1. $|\mu_f(x)'| \leq (y_2(\beta_m) - y_1(\alpha_m))^2 \; \forall x \in [\alpha_m, \beta_m]$, *if f, y_1, and y_2 are increasing over $[\alpha_m, \beta_m]$.*
2. $|\mu_f(x)'| \leq (y_2(\alpha_m) - y_1(\beta_m))^2 \; \forall x \in [\alpha_m, \beta_m]$, *if f, y_1, and y_2 are decreasing over $[\alpha_m, \beta_m]$.*

PROOF. The computations are identical to those for Lemma 13.1. Type **id** segments satisfy the sandwich inequalities $y_1(x) \le f(x) \le y_2(x)$, $\forall\, x \in [\alpha_m, \beta_m]$, if all three functions are increasing and $y_2(x) \le f(x) \le y_1(x)$, $\forall\, x \in [\alpha_m, \beta_m]$, if all three functions are decreasing, comp. the proof of Theorem 6. This completes the argument. ∎

Iterated function applications $g_{f,m}(x) = x^{(1)}, g_{f,m} \circ g_{f,m}(x) = x^{(2)}, \ldots$ may eventually leave the relevant kernel $rk_f([a_m, b_m])$ for some $x \in rk_f([a_m, b_m])$. This is agreeable with the notion of *explored* subintervals (comp. Definition 3) of relevant segments or relevant kernels of type **id**.

LEMMA 24. *Let* $rk_f([a_m, b_m]) =: [\alpha'_m, \beta'_m]$ *be a relevant kernel of type* **id**, *where* f, y_1, *and* y_2 *are increasing.*

1. *If* $x^{(k)} < \alpha'_m$ *for some* $k \in \mathbb{N}$, *then* $g_{f,m}(z), g_{f,m} \circ g_{f,m}(z), \ldots$ *eventually leaves* $rk_f([a_m, b_m])$ *for all* $z \in [\alpha'_m, x^{(1)}]$.
2. *If* $x^{(k)} > \beta'_m$ *for some* $k \in \mathbb{N}$, *then* $g_{f,m}(z), g_{f,m} \circ g_{f,m}(z), \ldots$ *eventually leaves* $rk_f([a_m, b_m])$ *for all* $z \in [x^{(1)}, \beta'_m]$.

This allows to apply the global procedure **GL1** for all monotone segments where necessary resulting in a global quantization procedure for nonmonotone functions f.

GL3

1. Initialization. For each monotonicity segment $[a_m, b_m]$ the relevant segment $r_m = [\alpha_m, \beta_m]$ is computed, $m = 1, \ldots, M$. Set $L \leftarrow \emptyset$.
2. Iteration. Each relevant segment $[\alpha_m, \beta_m]$ where $\operatorname{sgn} f'(x) = \operatorname{sgn}(y_2(x) - y_1(x))$ is satisfiable is searched for local minima of μ_f.
   (a) If $[\alpha_m, \beta_m]$ is of type **diff**, then it is tested for containing a fixed point of $g_{f,m}$ according to $(y_1(\alpha_m) + y_2(\alpha_m) - 2f(\alpha_m))(y_1(\beta_m) + y_2(\beta_m) - 2f(\beta_m)) \le 0$. If so, the fixed point x_0 is computed and $L \leftarrow L \cup \{x_0\}$.
   (b) If $[\alpha_m, \beta_m]$ is of type **id**, then the relevant kernel $rk_f([a_m, b_m]) =: [\alpha'_m, \beta'_m]$ is computed. If not void, it is searched for by **GL1** resulting in a finite list X_0 containing all local minima of μ_f over $[\alpha_m, \beta_m]$. $L \leftarrow L \cup X_0$.
3. $\arg\min_{z \in L} \mu_f(z) = x_2^0$.

4.2.2. Two and more levels

Local minima of the quantization objective need not be comparable in the component-wise sense for nonmonotone functions even in the three-level case.

EXAMPLE 6. Three-level quantization is considered. Therefore, the step function $f^* : [0, 6] \to [0, 3]$

$$f^*(x) = \begin{cases} 1, & \text{if } 0 \le x < 1, \\ 3, & \text{if } 1 \le x < 2, \\ 0, & \text{if } 2 \le x < 4, \\ 3, & \text{if } 4 \le x < 5, \\ 1, & \text{if } 5 \le x \le 6, \end{cases}$$

is deformed to be a continuous function $f : [0, 6] \to [0, 3]$ with parameter $0 < \varepsilon < 1/6$ such that $f(1) = f(5) = 5/4 - \varepsilon/2$, $f(2) = f(4) = 1 - \varepsilon/2$,

$$\int_0^1 f(u)\,du = \int_5^6 f(u)\,du = 1 + \varepsilon, \quad \int_1^2 f(u)\,du = \int_4^5 f(u)\,du = 3 - 5\,\varepsilon, \quad \int_2^4 f(u)\,du = 2\,\varepsilon.$$

The quantization objective $\mu_f(x_2, x_3)$ has local minima at $(1, 5)$ and $(2, 4)$ which are incomparable in the component-wise order. ∎

Example 6 further reveals that the interlacing result for local minima of the quantization objective is not true for nonmonotone functions.

The fixed-point function in the multiple level nonmonotone case is given by

$$g_f(x) := \begin{pmatrix} g_{f,m(2)}(x) \\ \vdots \\ g_{f,m(n-1)}(x) \end{pmatrix} := \begin{pmatrix} f_{m(2)}^{-1}\left(\frac{(y(x_1,x_2)+y(x_2,x_3))}{2}\right) \\ \vdots \\ f_{m(n-1)}^{-1}\left(\frac{(y(x_{n-2},x_{n-1})+y(x_{n-1},x_n))}{2}\right) \end{pmatrix},$$

with $x_i \in [a_{m(i)}, b_{m(i)}]$.

The fixed-point function need not to be defined for all $x \in \Delta([a,b])$. The next result serves this analysis.

LEMMA 25. *Let $[a_m, b_m]$ be a monotonicity segment of f and let values $x_i, x_{i+1} \in [a_m, b_m]$ be fixed with $x_i < x_{i+1}$. Then*

1. *$y(x_i, \cdot)$ is increasing over $[x_i, b_m]$ and $y(\cdot, x_{i+1})$ is increasing over $[a_m, x_{i+1}]$ if f is increasing over $[a_m, b_m]$,*
2. *$y(x_i, \cdot)$ is decreasing over $[x_i, b_m]$ and $y(\cdot, x_{i+1})$ is decreasing over $[a_m, x_{i+1}]$ if f is decreasing over $[a_m, b_m]$.*

LEMMA 26. *At most the two outer of the coordinates $x_i < x_{i+1} < \cdots < x_j$ which lie in the same monotonicity segment have no image under g_f.*

PROOF. Let $x_i, \ldots, x_j \in [a_m, b_m]$ and suppose f is increasing over $[a_m, b_m]$. According to Lemma 25 all x_k with $k \in \{i+1, \ldots, j-1\}$ then satisfy

$$\begin{aligned} l_m \le f(x_{k-1}) &\le \frac{f(x_{k-1})+f(x_k)}{2} \le \frac{y(x_{k-1},x_k)+y(x_k,x_{k+1})}{2} \\ &\le \frac{f(x_k)+f(x_{k+1})}{2} \le f(x_{k+1}) \le u_m. \end{aligned}$$

Hence, f_m^{-1} is defined for $(y(x_{k-1},x_k)+y(x_k,x_{k+1}))/2$. ∎

Trivially, each nonvoid relevant kernel contains a local minimum of the quantization objective in case of only two levels. A similar result holds in a more general sense.

THEOREM 7. *Each nonvoid relevant kernel $rk_f([a_m, b_m])$ of type* **id** *where both functions y_1 and y_2 intersect with f contains all coordinates of a local minimum $(x_{2,0}, \ldots, x_{n-1,0})$ of μ_f.*

PROOF. SKETCH. Function f is assumed to be increasing over $rk_f([a_m, b_m]) = [\alpha'_m, \beta'_m]$. Then $y_1(x) \le f(x) \le y_2(x)$, $\forall x \in [\alpha'_m, \beta'_m]$ and all level functions are increasing in $x_2, \ldots, x_{n-1}$. The sequence $x^{(1)}, x^{(2)} = g_f(x^{(1)}), \ldots$ can thus be shown to converge to a fixed point of g_f with all coordinates in $[\alpha'_m, \beta'_m]$ if initialized with $x^{(1)} = (\alpha'_m, \ldots, \alpha'_m) \in \mathbb{R}^{n-2}$. The sequence $(x^{(k)})_{k=1}^{\infty}$ is increasing in the componentwise order, i.e., $x^{(k)} \le x^{(k+1)}$, $\forall k$. ∎

Let $x_{i,0}$ be a coordinate of a local minimum of the quantization objective $\mu_f(x_2, \ldots, x_{n-1})$ and let $x_{i,0} \in [a_m, b_m]$. Then $x_{i,0}$ lies in a set of the form

$$\begin{aligned} rk_f([a_m,b_m]; x_{i-1}, x_{i+1}) := \Big\{ x \mid x \in [a_m,b_m] \cap [x_{i-1}, x_{i+1}] \text{ with } l_m \\ \le \frac{y(x_{i-1},x)+y(x,x_{i+1})}{2} \le u_m \text{ and } \min\{y(x_{i-1},x), y(x,x_{i+1})\} \\ \le f(x) \le \max\{y(x_{i-1},x), y(x,x_{i+1})\} \Big\}, \end{aligned}$$

with suitable values $x_{i-1} < x_{i+1}$. The set $rk_f([a_m, b_m]; x_{i-1}, x_{i+1})$ is called extended relevant kernel and it may be larger, smaller, or incomparable to the relevant kernel $rk_f([a_m, b_m])$. Choosing $x_{i-1}, x_{i+1} \in [a_m, b_m]$ with $x_{i-1} < x_{i+1}$ and $x_{i+1} - x_{i-1}$ sufficiently small always results in the extended relevant kernel being not empty, $rk_f([a_m, b_m]; x_{i-1}, x_{i+1}) \neq \emptyset$. Theorem 6 applies here too and each extended relevant kernel either is of type **id** or **diff**.

LEMMA 27. *An extended type* **diff** *kernel with respect to coordinates of a local minimum of the quantization objective contains at most one coordinate of that local minimum.*

PROOF. Obvious. ∎

For saw-tooth like functions, i.e., functions with $l_1 = \cdots = l_M$ and $u_1 = \cdots = u_M$ the global minimum of $\mu_f(x_2, \ldots, x_{n-1})$ can be computed by **GL2 large** for all $n \geq 3$. Thereby, each coordinate of successice approximations is kept within the same relevant kernel, different coordinates possibly belonging to different kernels depending on the initial value. In general, the procedure **GL2 large** will be modified by choosing the inverse of f over a suitable monotonicity segment which intersects with the corresponding projection of the current search area. Suppose that for some $x^{(k)} \in [c, d]$ and coordinate i with $x_i^{(k)} \in [a_{m(i)}, b_{m(i)}]$ the i^{th} coordinate of $g_f(x^{(k)})$ is undefined, i.e., $(y(x_{i-1}^{(k)}, x_i^{(k)}) + y(x_i^{(k)}, x_{i+1}^{(k)}))/2 \notin [l_{m(i)}, u_{m(i)}]$. The coordinate i of $g_f(x^{(k)})$ is defined, if the inverse of f is taken over another, suitable monotonicity segment $[a_{m^*(i)}, b_{m^*(i)}]$. The transition from one monotonicity segment to another is only performed if g_f where undefined otherwise and the transition is done so that $g_f(x^{(k)})$ stays within a search interval $[c, d]$ whenever possible. A global solution of the quantization problem can then be obtained in principle by the following procedure.

GL4

1. Initialization. The vectors $x^{(1)} = (a, \ldots, a)$ and $x^{(1)} = (b, \ldots, b)$ are chosen as initial values for $x^{(k+1)} = g_f(x^{(k)})$ resulting in the limits z_1 and z_2, respectively. Consideration is restricted to the interval $[z_1, z_2]$ which is labeled *explored* if and only if $z_1 = z_2$. $F \leftarrow \min\{\mu_f(z_1), \mu_f(z_2)\}$, $Li \leftarrow \{\arg\min\{\mu_f(z_1), \mu_f(z_2)\}\}$.
2. Iteration. As long as there is a subinterval $[c, d]$ of $[z_1, z_2]$ without a label *explored* do
   (a) Initialization of **FP**. A vector $x^{(1)} \in \text{conv}\{c, d\}$ is chosen as initial value for the successive approximations of **FP**. $\lim_{k\to\infty} x^{(k)} = z$.
   Termination analysis of **FP**.
   i. If $\lim_{k\to\infty} x^{(k)} = z \in [c, d]$ and if z was reached as limit (with other initial value $x^{(1)}$) before, then $[c, d]$ is devided by a hyperplane orthogonal to one axis and through $x^{(1)}$.
   ii. If $\lim_{k\to\infty} x^{(k)} = z \in [c, d]$ and if z was not reached as limit before, then $[c, d]$ is devided by a hyperplane orthogonal to one axis and through z. If $f(z) < F$, then $F \leftarrow f(z)$ and $Li \leftarrow \{z\}$. If $f(z) = F$, then $Li \leftarrow Li \cup \{z\}$.
   iii. If $x^{(k+1)} \notin [c, d]$ but $x^{(1)}, \ldots, x^{(k)} \in [c, d]$ and if each earlier obtained limit $z_0 \in [c, d]$ lies on an earlier established hyperplane, then $[c, d]$ is devided by a hyperplane orthogonal to one axis and through $x^{(1)}$.
   iv. If $x^{(k+1)} \notin [c, d]$ but $x^{(1)}, \ldots, x^{(k)} \in [c, d]$ and if an earlier obtained limit $z_0 \in [c, d]$ does not lie on any earlier established hyperplane, then $[c, d]$ is devided by a hyperplane orthogonal to one axis and through z_0.
   (b) Intervals in $[z_1, z_2]$ without label *explored* are tested for receiving this label.

5. DEFORMATION AND HEURISTIC SEARCH

Initializations of heuristic fixed-point routines as well as approximations of optimal quantizations can be computed by suitable deformations of f.

5.1. Monotonicity vs. Nonmonotonicity

A monotone substitute for a nonmonotone function f is its total variation $\text{Var}(f)_a^x$. The optimal quantization of $\text{Var}(f)_a^x$ can be computed exactly by **GL1**, **GL2 small**, and **GL2 large**, respectively. It provides an approximately optimal quantization of function f.

5.2. Convexity vs. Nonconvexity

Initial solutions to the quantization problem of a monotone function f which is neither convex nor concave can be constructed by a convex lower estimate f_l and a concave upper estimate f_u

$$f_l := \sup\{\gamma \mid \gamma \text{ is convex and } \gamma(x) \leq f(x)\ \forall x \in [a,b]\}, \text{ and}$$
$$f_u := \inf\{\gamma \mid \gamma \text{ is concave and } \gamma(x) \geq f(x)\ \forall x \in [a,b]\}.$$

By construction, function f_l is convex and function f_u is concave. f_l is also called convex hull of f (as it corresponds to the convex hull of the epigraph of f), which is the greatest convex function below f, see e.g., [17, p. 36]. Monotonicity of f carries over to both f_l and f_u. The quantization objectives of f_l and f_u tend to have fewer extrema than that of f. The partition of either of the two optimal quantizations of f_l and f_u can be chosen as initialization of the procedure **GenP** for the original function f.

If f_l and f_u are computationally too difficult to obtain, any convex lower monotone bound or concave upper monotone bound can be used in principle.

An initial partition for the quantization problem can also be given without resorting to convexity or concavity. Let functions A and B denote half of the area between an increasing function f and the horizontal lines y_1 and y_2. More precisely

$$A(x) := \int_a^{f^{-1}(y_1(x))} y_1(x) - f(u)\,du = \int_{f^{-1}(y_1(x))}^{x} f(u) - y_1(x)\,du,$$
$$B(x) := \int_x^{f^{-1}(y_2(x))} y_2(x) - f(u)\,du = \int_{f^{-1}(y_2(x))}^{b} f(u) - y_2(x)\,du.$$

Both $A(x)$ and $B(x)$ are well defined as each of the integrals for A, respectively, B are equal, comp. the balancing arguments in Section 3.

LEMMA 28. *Let f be increasing. Function A is then increasing while B is decreasing.*

PROOF. Only for $A(x)$. Let $x' > x$. Then

$$A(x') = \int_a^{f^{-1}(y_1(x'))} y_1(x') - f(u)\,du \geq \int_a^{f^{-1}(y_1(x))} y_1(x') - f(u)\,du$$
$$\geq \int_a^{f^{-1}(y_1(x))} y_1(x) - f(u)\,du = A(x). \qquad \blacksquare$$

Among the local minima of μ the global minimum may in some cases be close to a point, where the areas indicated by A and B are equal. A zero x_0 of the continuous, increasing function $A - B$ can be approximately found e.g., by the *regula falsi*, since

$$A(a) - B(a) < 0 < A(b) - B(b).$$

The fixed-point procedure **FP** can be initialized with x_0. A zero of $A - B$ is generally not a local minimum of μ even for convex f and it may be a local maximum in the worst case.

EXAMPLE 7. CONTINUATION OF EXAMPLE 2. For f as in Example 2, the bound f_l is given by

$$f_l(x) = \begin{cases} x, & \text{if } 0 \leq x \leq 11, \\ 10x - 99, & \text{if } 11 \leq x \leq 12. \end{cases}$$

The unique optimal quantization of f_l is given by $x_l^0 = 11$, where x_l^0 is not a zero of the respective function $A_l - B_l$: $A_l(11) - B_l(11) = 121/8 - 5/4 > 0$. x_l^0 leads to one of the global minima of the quantization of f when used as initial value of the fixed-point procedure.

However, symmetry of the original function f implies $A(6) - B(6) = 0$. The zero $\bar{x} = 6$ of $A - B$ coincides with a local maximum of the quantization objective μ. The zero $\bar{x} = 6$ is the unique local maximum in the interior of the domain of f. $\blacksquare$

5.3. Heuristics

In case of $n-1=2$ levels, function f has a closely related linear surrogate depending on the current break point.

LEMMA 29. *For each f and fixed $x_0 \in (a,b)$ the piecewise linear surrogate function*

$$f_{\text{lin}}(x) = \begin{cases} f(x_0) + 2\,(y_2(x_0) - f(x_0)) \cdot \dfrac{(x-x_0)}{(b-x_0)}, & \text{if } x_0 \le x \le b, \\ f(x_0) + 2\,(f(x_0) - y_2(x_0)) \cdot \dfrac{(x-x_0)}{(x_0-a)}, & \text{if } a \le x \le x_0, \end{cases}$$

has the same levels as f meaning that

$$\int_a^{x_0} f_{\text{lin}}(x)\,dx = \int_a^{x_0} f(x)\,dx \qquad \text{and} \qquad \int_{x_0}^{b} f_{\text{lin}}(x)\,dx = \int_{x_0}^{b} f(x)\,dx.$$

The surrogate function f_{lin} need not be monotone but its optimal quantization can be computed analytically.

LEMMA 30. *The optimal quantization function s_3^0 for f_{lin} has break point x_2^0 specifiable by a closed formula.*

PROOF. The optimal break point $x = x_2^0$ is a solution of $y_1(x) + y_2(x) = 2f(x)$ and $\operatorname{sgn} f'(x) = \operatorname{sgn}(y_2(x) - y_1(x))$. The first equation amounts to solving the cubic equation (comp. the proof of Theorem 3) $F(x)(b-x) + (F(b) - F(x))(x-a) = 2f(x)(x-a)(b-x)$ with two known solutions $x = a$ and $x = b$. The candidates from both linear segments are tested for global optimality by computing their μ_f values. ∎

This allows a heuristic procedure for approximating the optimal two-level quantization of a nonmonotone function f with arbitrary curvature.

H1

1. Initialization. An arbitrary break point $x_2^{(1)} \in (a,b)$ is selected. Set $k \leftarrow 1$.
2. Iteration. Repetition until some stopping criterion is met:
   (a) Computation of levels $y_1(x_2^{(k)}) = y(a, x_2^{(k)})$ and $y_2(x_2^{(k)}) = y(x_2^{(k)}, b)$.
   (b) Computation of a piecewise linear surrogate function

$$f_{\text{lin}}(x) = \begin{cases} f\left(x_2^{(k)}\right) + 2\left(y_2\left(x_2^{(k)}\right) - f\left(x_2^{(k)}\right)\right) \cdot \dfrac{\left(x - x_2^{(k)}\right)}{\left(b - x_2^{(k)}\right)}, & \text{if } x_2^{(k)} \le x \le b, \\ f\left(x_2^{(k)}\right) + 2\left(f\left(x_2^{(k)}\right) - y_2\left(x_2^{(k)}\right)\right) \cdot \dfrac{\left(x - x_2^{(k)}\right)}{\left(x_2^{(k)} - a\right)}, & \text{if } a \le x \le x_2^{(k)}. \end{cases}$$

   (c) Computation of the optimal quantization of f_{lin} with break point $x_2^{(k+1)}$.
   (d) $k \leftarrow k+1$.

A greedy heuristic without a surrogate function can be constructed by always choosing the locally best branch of f.

H2

1. Initialization. An arbitrary break point $x_2^{(1)} \in (a,b)$ is selected. Set $k \leftarrow 1$.
2. Iteration. Repetition until some stopping criterion is met:
   (a) Computation of levels $y_1(x_2^{(k)}) = y(a, x_2^{(k)})$ and $y_2(x_2^{(k)}) = y(x_2^{(k)}, b)$.

(b) Computation of new candidate partitions $x_{2,m}^{(k+1)} = g_{f,m}(x_2^{(k)}) = f_m^{-1}((y_1^{(k)} + y_2^{(k)})/2)$, where $x_{2,m}^{(k+1)}$ is computed for all $m \in \{1, \ldots, M\}$ with $(y_1^{(k)} + y_2^{(k)})/2 \in [l_m, u_m]$ and valid sign condition $\operatorname{sgn} f'(x_{2,m}^{(k+1)}) = \operatorname{sgn}(y_2^{(k)} - y_1^{(k)})$.

(c) The best of the candidates from (b) is chosen according to $x_2^{(k+1)} = \arg\min_m \mu_f(x_{2,m}^{(k+1)})$.

(c) $k \leftarrow k + 1$.

There is at least one $x_{2,m}^{(k+1)}$ in each iteration of Step 2(b) as otherwise level $y_1^{(k)}$ or $y_2^{(k)}$ could be shown to be not optimal for the current break point $x_2^{(k)}$. Numerous variants of **H2** can be designed. One such variant is to simplify the computations of Step 2(c) by fixing the levels so that $x_2^{(k+1)}$ is chosen as the best from the $x_{2,m}^{(k+1)}$ according to Lemma 8. Another variant is to restrict the search for the next break point to a single branch of f in each iteration and to omit Step 2(c). The current branch f_m is abandoned only if $(y_1^{(k)} + y_2^{(k)})/2 \notin [l_m, u_m]$. The new branch may be an arbitrary one satisfying the fixed point and sign conditions.

For the number of levels being a power of 2, i.e., $n - 1 = 2^\nu$, the optimal quantization can be approximated by subdivisioning and by local optimization.

H3

1. Initialization. The 2 level problem is solved to optimality by **GL3**. Set $k \leftarrow 1$.
2. Iteration. Repetition until $k = 2^\nu$:
   (a) A value is chosen from each of the intervals $(a, x_2), (x_2, x_3), \ldots, (x_{2^k}, b)$.
   (b) The new partition of $[a, b]$ is chosen as inital partition to **FP** to result in $a < x_2 < \cdots < x_{2^{k+1}} < b$.
   (c) $k \leftarrow 2k$.

6. LIMIT BEHAVIOR

Stability or limit considerations are motivated by quantizations possibly converging to function f as the number of break points increases unboundedly and for a fixed number of break points by a sequence of functions $(f_i)_{i=1}^\infty$ converging to f.

LEMMA 31. *Let $f \in C(D)$ with $D = [a, b]$ and $1 \le p < \infty$. Then $\|f - s_n^0\|_p \to 0$ $(n \to \infty)$.*

PROOF. Function f can be assumed to be nonnegative. Thus, it can be approximated pointwise on D by a sequence of step functions $\sigma_n \in S_n$ with values not greater than those of f; see for example [18, p. 63]. The dominated convergence theorem then implies $\|f - \sigma_n\|_p \to 0$ $(n \to \infty)$ as all step functions and their bound are integrable. Minimality of s_n^0 leads to the desired convergence. ∎

The rate of convergence of $\|f - s_n^0\|_p$ is at least linear. This is shown here for monotone functions only.

THEOREM 8. *For monotone f and $1 \le p < \infty$ the optimal quantization is bounded by*

$$\left\|f - s_n^0\right\|_p \le \frac{\operatorname{Var}(f)_a^b}{2(n-1)} (b-a)^{1/p} = O\left(n^{-1}\right), \qquad (n \to \infty).$$

PROOF. Let s_n^* be the step function from S_n with break points and levels equally deviding the total variation of an increasing function f

$$x_i^* := (\operatorname{Var}(f)_a^\cdot)^{-1}\left(f(a) + (i-1)\frac{\operatorname{Var}(f)_a^b}{n-1}\right), \qquad i = 2, \ldots, n-1$$

$$y_i^* := f(a) + (2i-1)\frac{\operatorname{Var}(f)_a^b}{2(n-1)}, \qquad i = 1, \ldots, n-1.$$

Then

$$\|f - s_n^0\|_p^p \leq \|f - s_n^*\|_p^p = \sum_{i=1}^{n-1} \int_{x_i^*}^{x_{i+1}^*} (f(x) - y_i^*)^p \, dx$$
$$\leq \sum_{i=1}^{n-1} (x_{i+1}^* - x_i^*) \left(\frac{\mathrm{Var}(f)_a^b}{2(n-1)} \right)^p = (b-a) \left(\frac{\mathrm{Var}(f)_a^b}{2(n-1)} \right)^p . \quad \blacksquare$$

LEMMA 32. *Convergence of a sequence $(f_i)_{i=1}^{\infty}$ carries over to suitable optimal quantizations $s_n^0(i)$ for fixed number n of break points, i.e., $f_i \to f \Longrightarrow s_n^0(i) \to s_n^0$ $(i \to \infty)$, where s_n^0 is a suitable optimal quantization of f.*

PROOF. SKETCH. The result follows from the continuity of μ_f in changes of f. $\blacksquare$

In case f_i or f have alternate optimal quantizations convergence only holds for particular quantizations and f may have an optimal quantization which is not the limit of any sequence of optimal quantizations of the f_i.

7. CONCLUSION AND OUTLOOK

An approach has been presented which has the potential to solve the quantization problem for normal functions to global optimality. For particular functions such as polynomials and splines, particular versions of the approach and closed form solutions for low degree cases are to be developed. A couple of technical issues are also left to future work as it is conjectured that for any increasing and sufficiently smooth function $g : [a, b] \to (a, b)$ with only finite many fixed points there is a normal increasing function f on $[a, b]$ such that $g(x) = g_f(x) \; \forall x \in [a, b]$. Another conjecture is that the number of local minima of the objective $\mu_f(x_2, \ldots, x_{n-1})$ is decreasing in n for monotone functions f.

Additional constraints can be added to the quantization problem. These include the restriction that all break points lie in a given connected or disconnected subset of D or that the step function must lie above or below the given function f. In the setting of stochastic processes whose paths are almost surely continuous, optimal quantization is to be investigated for providing a maximum likelihood estimate by jump processes.

REFERENCES

1. A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer, Boston, MA, (1992).
2. D. Mumford and J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, *Communications on Pure and Applied Mathematics* **42**, 577–685, (1989).
3. D. Mumford, *Pattern Theory: A Unifying Perspective*, (Preprint), Harvard University, (1993).
4. U. Grenander, *General Pattern Theory*, Clarendon, Oxford, (1993).
5. H.-H. Chen, Y.-S. Chen and W.-H. Hsu, Low rate sequence image coding via vector quantization, *Signal Processing* **26**, 265–283, (1992).
6. K. Donner, Hierarchical approximation for pattern recognition, In *Proceedings Applications of Artificial Intelligence (AAAI 1991)*, Prague, pp. 213–226, (1991).
7. W. Koller, *Segmentierung und Diagnose von Gefäßen mit Atlasunterstützung*, VDI-Verlag, Düsseldorf, (1995).
8. H. Schmidt, Histogrammbasierte Klassifikation von Sensordaten, Dissertation, Universität Ulm, (1993).
9. A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge MA, (1979).
10. J.W. Schmidt, Constrained smoothing of histograms by quadratic splines, *Computing* **48**, 97–107, (1992).
11. G. Koeppler, C. Lopez and J.M. Morel, A multiscale algorithm for image segmentation by variational methods, *SIAM Journal on Numerical Analysis* **31**, 282–299, (1994).
12. G. DalMaso, J.M. Morel and S. Solimini, A variational method in image segmentation: Existence and approximation results, *Acta Mathematica* **168**, 89–151, (1992).
13. D.W. Scott, *Multivariate Density Estimation*, Wiley, New York, (1992).
14. E.W. Cheney, *Introduction to Approximation Theory*, McGraw-Hill, New York, (1966).
15. R. Horst and H. Tuy, *Global Optimization*, Springer, Berlin, (1990).
16. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze der Analysis*, Band 2, Springer, Berlin, (1971).
17. R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, (1970).
18. H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, de Gruyter, Berlin, (1974).